Università di Pisa

Dipartimento di Fisica
Corso di Laurea Magistrale in Fisica

Tesi di Laurea Magistrale

Study of the impact of the planned CMS detector upgrade on the Higgs to b-quarks measurements

Candidato
Valeria Botta

Relatori
Prof. A. Messineo
Dr. A. Rizzi

A.A. 2014/2015

"The only thing greater
than the power of the mind
is the courage of the heart."

John Nash

Alla mia famiglia

A Thomas

Abstract

The aim of this Thesis is to study the Higgs boson coupling to bottom quarks in the next Runs of the Large Hadron Collider (LHC). In particular the performance of a simplified analysis in the context of the CMS experiment is studied in different scenarios, starting from the upcoming LHC Run 2 and extending to long term runs, after upgrades of the LHC machine and of the CMS detector. This Thesis work also includes contribution to the qualification of silicon pixel modules for the new CMS pixel tracker that will replace the current one at the end of 2016.

For a Higgs boson of 125 GeV, the decay to a bottom quark pair ($b\bar{b}$) is the dominant mode, with a branching ratio of approximately 58%. However, the prevailing QCD multijet background events prevent an inclusive search of the Higgs boson in the $b\bar{b}$ decay channel. Therefore searches of the H($b\bar{b}$) decay at the LHC consider particular production modes of the Higgs boson: the vector boson fusion and the associated production with a vector boson or with a top quark pair.

In this Thesis, the associated production of the Higgs with a Z boson has been studied, when the latter decays into a pair of opposite sign muons or electrons (referred to as leptons hereafter). In such final state the signature of two isolated opposite sign leptons from the Z boson decay helps for triggering and rejecting background events. Due to the nearly hermetic coverage of the CMS detector and thanks to its silicon tracker, its high granularity electromagnetic calorimeter and its excellent muon system, leptons originating from Z bosons can be efficiently identified and precisely reconstructed.

The reconstruction of the Higgs boson in the $b\bar{b}$ final state relies on the identification of b-jets, originating from the hadronization of b-quarks into B hadrons. Therefore, dedicated b-tagging algorithms have been developed to identify b-jets, being based on distinguishing features of the B hadrons. The latter are characterized by relatively long lifetimes, such that they decay inside the tracker volume and tracks of daughter charged particles identify a secondary displaced vertex within the jet. Moreover, such tracks also tend to have larger impact parameters. A good ability in b-jet identification relies on efficient and high quality tracking and vertexing, that can only be achieved with an appropriate performance of a pixel tracking detector.

Therefore, the second main topic of this Thesis deals with activities that contribute to upgrade the CMS silicon pixel tracker to ensure and improve its performance. The current

CMS silicon pixel tracker was designed to sustain a maximum fluence of $1.6 \times 10^{15} \mathrm{n}_{eq}/\mathrm{cm}^2$ and to operate at a maximum instantaneous luminosity of 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. As the latter limits will be exceeded at the end of the LHC Run 2 in 2016, the pixel tracker must be replaced with a new upgraded one by then. The so called Phase 1 upgrade pixel tracker is currently being built and is expected to be installed at the end of 2016. It will have one additional silicon layer with respect to the current one and the innermost layer closer to the interaction point, without increasing the amount of material. Such improvements will result in a higher tracking efficiency, which is important for tracking performance and hence for b-jet identification. Additional extended upgrades of the CMS tracker, usually referred to as Phase 2 upgrades, are foreseen around 2023 before the beginning of High Luminosity Phase of the LHC, needed to retain the same excellent detector performances in an even more challenging environment.

This Thesis is organized in seven Chapters. **Chapter 1** introduces the Higgs-Brout-Englert mechanism of spontaneous symmetry breaking in the Standard Model of Particle Physics and summarizes the phenomenology of Higgs boson production and decays at hadron colliders. Latest experimental searches for the Higgs boson at the LHC leading to its discovery in 2012 are also cited.
In **Chapter 2** the layout and operation of the LHC machine are briefly summarized. Then the subsystems of the CMS detector are described, including the main upgrades planned for the coming years, highlighting the improvements in the inner and outer silicon tracking detector.
Chapter 3 focuses on the building and commissioning of the Phase 1 upgrade pixel tracker, that shall be installed at the end of 2016. After illustrating the components of pixel barrel modules, activities currently going on in the Pisa production center are described. In **Chapter 4** as an introduction to the analysis work the main strategies and algorithms for particle identification and reconstruction adopted by the CMS collaboration are presented.
The first Section of **Chapter 5** summarizes the strategy and results for the search of the Higgs boson produced in association with a vector boson and decaying to a $\mathrm{b\bar{b}}$ pair, performed on data collected with the CMS detector until the end of 2012. The second Section describes the preparation of the analysis for the LHC run at a center of mass energy of 13 TeV. First, some basic kinematics studies comparing Monte Carlo samples at 8 and 13 TeV are performed. Then, the performance of a simple cut-based invariant mass analysis is evaluated and the choice of different b-tagging discriminators and working points is evaluated. In **Chapter 6** the impact of the b-tagging performance on the $\mathrm{Z}(ll)\mathrm{H}(\mathrm{b\bar{b}})$ sensitivity is estimated in future detector upgrade scenarios. Due to the impossibility of producing large full simulation samples for the most important background processes in such future scenarios, the studies are based on a two-dimensional parametrization of the b-tagging efficiency, derived from fully simulated $\mathrm{t\bar{t}}$ samples, available for all the considered scenarios. Such parametrization is then used to estimate the sensitivity of the $\mathrm{Z}(ll)\mathrm{H}(\mathrm{b\bar{b}})$ analysis in those future upgrade scenarios, highlighting the improvements that are expected to be achieved thanks to better tracking capabilities. In the last **Chapter 7** the most relevant results of this Thesis are summarized.

Contents

Chapter 1

The Standard Model and the Higgs Mechanism

The current understanding of high energy particle physics is summarized in the Standard Model (SM), describing the properties of known particles and their interactions. Since the 1970s when it has been developed, the Standard Model has successfully explained almost all experimental results spanning from discoveries of new particles, as the top quark [1][2] or the tau neutrino [3], to precise measurements of particle properties and couplings. The missing piece to complete the whole picture was the Higgs boson, predicted by the Brout-Englert-Higgs mechanism [4][5][6], that could explain how electroweak gauge bosons and fermions acquire mass. On July 4^{th}, 2012 the ATLAS [7] and CMS [8] collaborations announced the discovery of a new boson consistent with the Standard Model Higgs boson. Since then a detailed investigation program has started, in order to precisely measure the properties and couplings of the new boson and determine whether it behaves as predicted in the Standard Model or it is another kind of Higgs boson compatible with other theories beyond the Standard Model. On October 8^{th}, 2013 the Nobel prize in physics was awarded jointly to François Englert and Peter Higgs "for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider" [9]. However, the Standard Model shows the limitation of a model, because it is not including gravity and can not explain several experimental evidences, from the composition of dark matter to neutrino oscillations and the matter antimatter asymmetry.

In this Chapter, the mechanism of spontaneous symmetry breaking in the Standard Model is briefly presented. Such mechanism implies the existence of a new scalar particle, the Higgs boson. Therefore, processes of Higgs boson production at hadron colliders as well as the most important decay modes are discussed. Finally, the most relevant results obtained by the ATLAS and CMS Collaborations and leading to the Higgs boson discovery are summarized.

1.1 Particle Content of the Standard Model

The Standard Model (SM) of particle physics is a quantum non-abelian field theory, based on a SU(2) × U(1) × SU(3) gauge symmetry, where SU(2) × U(1) symmetry describes electromagnetic and weak interactions, unified in the Glashow-Weinberg-Salam theory [10][11][12], and the SU(3) symmetry is the basis for Quantum Chromodynamics (QCD), describing strong interactions.

According to the SM, matter is composed by spin 1/2 elementary particles classified in two groups, leptons and quarks. Each fermion has its own antiparticle, with the same mass but opposite quantum numbers. Leptons are organized into three doublets, each including a massive lepton and its corresponding neutrino

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \tag{1.1}$$

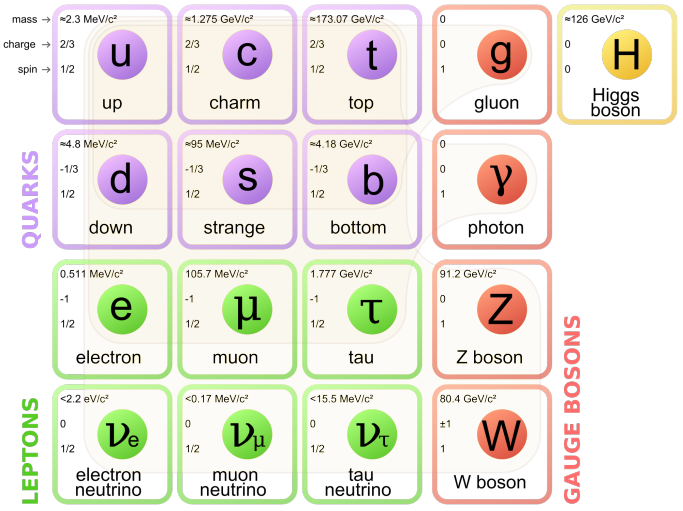
The lower components of the doublets has electromagnetic (EM) charge $Q = -1$ while neutrinos are neutral particles. Similarly, the quark sector has three isospin doublets

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} \tag{1.2}$$

with masses from a few MeV up to more than 170 GeV for the top quark [13]. Quarks have fractional EM charge, +2/3 and -1/3 for the quarks in the upper (up) and lower (down) row, respectively.

The SM predicts that forces are mediated by boson particles, the $W^{\pm}$ and Z bosons being responsible for the weak interaction, the photon (γ) for the EM force and gluons (g) for strong interactions. Figure 1.1 shows the SM particle content.

Figure 1.1 Particle content of the Standard Model: quarks, leptons and gauge bosons [14].



While the photon and the gluons are massless, the EW bosons and fermions have finite masses, whose values are experimentally measured. However, the direct introduction of a mass term in the Lagrangian would violate the gauge invariance on which the model is based. A solution to solve this puzzle has been proposed independently by Higgs, Brout and Englert in three papers published between 1963 and 1964 [5][4][6], and is commonly referred as the Higgs mechanism. Such solution is based on spontaneous symmetry breaking (SSB) introducing a new scalar field in the SM, that implies the existence of an additional scalar boson. In the following part of the Chapter, the basic idea of SSB in a gauge theory is presented, starting from simple examples and then generalizing to the electroweak sector of the Standard Model.

1.2 Spontaneous Symmetry Breaking in a Gauge Theory

A Simple Example In order to describe the main idea of symmetry breaking, the Lagrangian for a simple model with a real scalar field ϕ and a quartic potential is considered:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^2 - V(\phi) = \frac{1}{2}(\partial_\mu \phi)^2 - \frac{1}{2}\mu^2\phi^2 - \frac{1}{4}\lambda\phi^4 \tag{1.3}$$

The potential $V(\phi)$ has different shapes depending on the μ^2 sign:

- if $\mu^2 > 0$ the minimum of the potential is the state with $\phi = 0$
- if $\mu^2 < 0$ the minimum of the potential is located on the circumference of radius $v = \sqrt{-\mu^2/\lambda}$. Figure 1.2 shows the shape of the potential in this latter case.

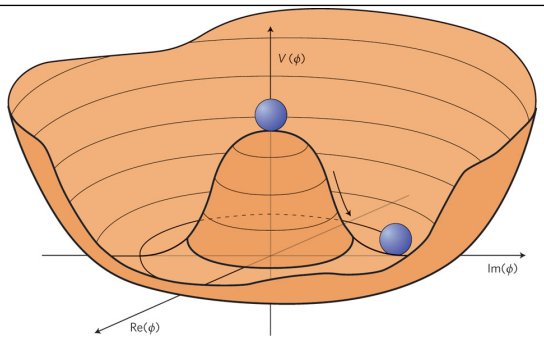
Only the latter case, having a vacuum expectation value (VEV) $v \neq 0$ allows spontaneous symmetry breaking. To investigate the particle spectrum of the model, the standard procedure is to expand the Lagrangian around its minimum. To do this, a shifted field $\eta = \phi - v$ is introduced and since $\partial_\mu v = 0$, neglecting constant terms and terms of order higher than η^2, the full Lagrangian becomes:

$$\begin{aligned} \mathcal{L}(\eta) &= \tfrac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta) + \tfrac{1}{2}\mu^2(\eta + v)^2 + \tfrac{1}{4}\lambda(\eta + v)^4 \\ &= \tfrac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta) - \lambda v^2 \eta^2 \end{aligned} \tag{1.4}$$

Although the Lagrangian in Equation 1.4 is still symmetric in ϕ, perturbations around the minimum are not, as $V(\eta) \neq V(-\eta)$: the symmetry has been spontaneously broken.

The same procedure of introducing a new scalar field with an appropriate potential and expanding the Lagrangian around the minimum can be applied to a global or local gauge invariant theory to introduce SSB.

Figure 1.2 The "mexican hat" shape of the quartic potential that allows for spontaneous symmetry breaking [15].



Breaking a local U(1) gauge symmetry To better figure out how a spontaneously broken symmetry allows for massive gauge bosons, in the following paragraph the SSB mechanism is applied to a theory based on a local U(1) symmetry. First, a complex scalar field $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$ with a quartic potential is added to the theory. Two new terms will then be added to the Lagrangian, without affecting the U(1) global symmetry:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi,\phi^\dagger) \tag{1.5}$$

where

$$V(\phi,\phi^\dagger) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2,\ \lambda > 0 \tag{1.6}$$

The covariant derivative is defined as $D_\mu = \partial_\mu - ieA_\mu$, A_μ being the field of the massless gauge boson, as in the EM theory, and the first term in Equation 1.5 representing the kinetic term.

In the interesting case $\mu^2 < 0$, the minimum for the potential (vacuum) is reached for all the values (ϕ_1, ϕ_2) satisfying the following relation:

$$\sqrt{{\phi_1}^2 + {\phi_2}^2} = v = \sqrt{-\frac{\mu^2}{\lambda}} \tag{1.7}$$

that describes a circumference of radius v in the (ϕ_1, ϕ_2) plane. After choosing one of the vacuum points $(\phi_1, \phi_2) = (v, 0)$ and defining shifted coordinates

$$\begin{aligned} \eta &= \phi_1 - v \\ \xi &= \phi_2 \end{aligned} \tag{1.8}$$

the vacuum state can be written as

$$\phi_0 = \frac{1}{\sqrt{2}}[(v+\eta) + i\xi] \tag{1.9}$$

Expanding the Lagrangian around the minimum defined in Equation 1.9 up to second order in the fields leads to the following expression:

$$\mathcal{L}(\eta,\xi) = \frac{1}{2}(\partial_\mu\eta\partial^\mu\eta) - \lambda v^2\eta^2 + \frac{1}{2}(\partial_\mu\xi\partial^\mu\xi) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}e^2v^2A_\mu A^\mu - evA_\mu(\partial^\mu\xi) + \text{inter. terms} \tag{1.10}$$

The physical meaning of the terms in Equation 1.10 can be easily understood. Two new particles have appeared as a consequence of the broken symmetry. The first two terms are the kinetic and mass terms for the η particle, while the third term is the kinetic energy of the massless particle ξ. The kinetic term for the photon field remains untouched but a mass-like term appears for the A_μ field: the previously massless gauge boson has now gained a mass $m_\mu = ev$, whose value is proportional to the VEV of the ϕ scalar field.

Although two new particles have apparently popped up in the theory, only one of them represent a physical particle. The ξ massless boson, known as the Goldstone boson, can be removed fixing a proper gauge. In a local gauge invariant theory, the field A_μ is in fact fixed up to a term $\partial_\mu\alpha$. The choice $\alpha = -\xi/v$ is such that all terms containing the ξ field no longer appear in the Lagrangian of Equation 1.10. This particular gauge is called *unitary* gauge.

In this representation, the ϕ field must be rotated accordingly by the phase α:

$$\phi_0^{'} = \phi_0\ e^{-i\xi/v} = e^{-i\xi/v}\frac{1}{\sqrt{2}}(v+\eta+i\xi) \equiv \frac{1}{\sqrt{2}}(v+h) \tag{1.11}$$

In the last step of Equation 1.11 the h real field has been introduced. The scalar part of the Lagrangian 1.5 written in the unitary gauge for the rotated field 1.11 is

$$\begin{aligned} \mathcal{L}_{scalar} &= (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi) \\ &= \tfrac{1}{2}(\partial_\mu h\partial^\mu h) - \lambda v^2h^2 + \tfrac{1}{2}e^2v^2A_\mu^2 + e^2vA_\mu^2h + \tfrac{1}{2}e^2A_\mu^2h^2 - \lambda vh^3 - \tfrac{1}{4}\lambda h^4 \end{aligned} \tag{1.12}$$

In the unitary gauge, the Lagrangian 1.12 of the local U(1) theory after breaking the symmetry contains a massive scalar particle h with mass $m_h = \sqrt{2\lambda v^2}$, a gauge massive field A_μ, the mixed A_μ, h terms describing the interaction of the Higgs field h with with the gauge field, and the Higgs self interaction.

1.3 The Higgs Mechanism in the Standard Model

The easiest choice [16] to break the local SU(2) × U(1) gauge symmetry of the Standard Model is adding an isospin doublet of complex scalar fields ϕ with weak hypercharge Y=1 to the theory:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.13}$$

The Lagrangian for the new field defined in 1.13 is

$$\mathcal{L}_{scalar} = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi) \tag{1.14}$$

where the potential has the form

$$V(\phi, \phi^\dagger) = \mu^2(\phi^\dagger \phi) + \lambda(\phi^\dagger \phi)^2, \text{ with } \lambda > 0,\ \mu^2 < 0 \tag{1.15}$$

and the covariant derivative is written as

$$D_\mu \equiv \partial_\mu + ig\frac{1}{2}\vec{\tau} \cdot \vec{W}_\mu + ig'\frac{1}{2}YB_\mu \tag{1.16}$$

where $\vec{\tau}$ are the Pauli matrices, B_μ and ${W_\mu}^i$, with $i = 1, 2, 3$ the fields of the massless vector gauge bosons. A possible convenient choice for the minimum of the Lagrangian is

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.17}$$

Substituting the latter vacuum expectation value ϕ_0 in the Lagrangian of Equation 1.14, gauge boson masses can be obtained expanding the term

$$\begin{aligned} &\left[\left(-ig\tfrac{\vec{\tau}}{2} \cdot \vec{W}_\mu - i\tfrac{g'}{2}B_\mu\right)\phi_0\right]^\dagger \left[\left(-ig\tfrac{\vec{\tau}}{2} \cdot \vec{W}_\mu - i\tfrac{g'}{2}B_\mu\right)\phi_0\right] = \\ &= (\tfrac{1}{2}vg)^2 W_\mu^+ W^{-\mu} + \tfrac{1}{8}v^2(W_\mu^3, B_\mu) \begin{pmatrix} g^2 & -gg' \\ -gg' & g'^2 \end{pmatrix} \begin{pmatrix} W^{3\mu} \\ B^\mu \end{pmatrix} \end{aligned} \tag{1.18}$$

where $W^\pm = (W^1 \mp iW^2)/\sqrt{2}$. From the first term of Equation 1.18 the W mass can be read to be

$$m_W = \frac{1}{2}vg \tag{1.19}$$

The second term in Equation 1.18 is not diagonal, meaning that the W^3_μ and B_μ field are not mass eigenstates. The mass matrix has two eigenvalues, corresponding to two eigenstates

$$\begin{aligned} A_\mu &= \frac{g'W^3_\mu + gB_\mu}{\sqrt{g^2+g'^2}} \quad \text{for} \quad m_\gamma = 0 \\ Z_\mu &= \frac{gW^3_\mu - g'B_\mu}{\sqrt{g^2+g'^2}} \quad \text{for} \quad m_Z = \tfrac{1}{2}v\sqrt{g^2+g'^2} \end{aligned} \tag{1.20}$$

In other words, the mass eigenstates A_μ and Z_μ are rotated of an angle θ_W with respect to the (W^3_μ, B_μ) basis:

$$\begin{aligned} A_\mu &= \cos\theta_W B_\mu + \sin\theta_W W^3_\mu \\ Z_\mu &= -\sin\theta_W B_\mu + \cos\theta_W W^3_\mu \end{aligned} \tag{1.21}$$

The W and Z bosons masses are related to the Weinberg angle θ_W:

$$\frac{m_W}{m_Z} = \cos\theta_W \tag{1.22}$$

The photon has instead remained massless, because the vacuum state ϕ_0 has been chosen to be neutral not to break the $\mathrm{U}(1)_{EM}$ symmetry, that is $\phi_0' = e^{i\alpha(x)Q}\phi_0 = \phi_0$.

As in the previous paragraph, the mass of the new scalar particle, the Higgs boson, is $m_h = \sqrt{2\lambda v^2}$. The value of the VEV is fixed and can be obtained knowing the Fermi constant and the muon mass, finding $v \approx 246$ GeV. However, since λ is a free parameter, no prediction on the Higgs boson mass is provided.

Fermion Masses When applying the Higgs mechanism to the full SM Lagrangian, terms of the form $-\lambda_f \bar{\psi}_L \phi \psi_R$ appear, where ψ is a fermionic spinor and the subscripts L, R refers to its left and right handed components. Due to the Higgs boson quantum numbers, such a term is invariant under a $\mathrm{SU}(2)\times\mathrm{U}(1)$ transformation. As for the gauge bosons, fermions acquire a finite mass due to the non-zero VEV of the ϕ field. As an example, the electron case is discussed. The relevant part of the Lagrangian is the following:

$$\begin{aligned} \mathcal{L}_e &= -\lambda_e \frac{1}{\sqrt{2}} \left[(\bar{\nu}, \bar{e})_L \begin{pmatrix} 0 \\ v+h \end{pmatrix} e_R + \bar{e}_R (0, v+h) \begin{pmatrix} \nu \\ e \end{pmatrix}_L \right] \\ &= -\frac{\lambda_e(v+h)}{\sqrt{2}} \left[\bar{e}_L e_R + \bar{e}_R e_L \right] \\ &= -\frac{\lambda_e v}{\sqrt{2}} \bar{e}e - \frac{\lambda_e}{\sqrt{2}} h \bar{e}e \end{aligned} \tag{1.23}$$

It includes a mass term and an interaction term between the electron and the Higgs boson. The coupling constant is the so called *Yukawa* coupling, and is a free parameter.

The mass of the electron

$$m_e = \frac{\lambda_e v}{\sqrt{2}} \tag{1.24}$$

is therefore not predicted, and the Yukawa coupling of the electron to the Higgs boson is proportional the electron mass itself. The electron example can then be generalized to the other fermions. However, from the Lagrangian in 1.23 it can be clearly seen that only down type fermion can gain mass through a Yukawa coupling. Due to this reason, a new gauge invariant term is added to the Lagrangian, again using the complex Higgs doublet:

$$\mathcal{L}_{up} = \bar{\chi}_L \tilde{\phi}^c \phi_R + \text{h.c.} \tag{1.25}$$

where $\tilde{\phi}^c$ is the representation of the Higgs field with hypercharge Y=-1 and is derived from the ϕ field as

$$\tilde{\phi}^c = -i\tau_2 \phi^* = -\frac{1}{\sqrt{2}} \begin{pmatrix} v+h \\ 0 \end{pmatrix} \tag{1.26}$$

The result is similar to what obtained in Equation 1.23, this time for the upper component of the ispospin doublet. The coupling with the Higgs boson is again proportional to the fermion mass.

1.4 Higgs Boson Phenomenology at Hadron Colliders

Higgs Boson Production Modes

At hadron colliders, there are several processes through which a Higgs boson can be produced [17]. In the following paragraphs, their features will be briefly discussed. Figure 1.3 shows the production cross sections for different processes at $\sqrt{s} = 8$ and 14 TeV, as a function of the Higgs boson mass [18].

Figure 1.3 Inclusive cross section for the production of a Standard Model Higgs boson from pp collisions at $\sqrt{s} = 8$ (top) and 14 TeV (bottom), as a function of the Higgs boson mass, separately for the different production modes. Results are obtained from theoretical calculations up to the order indicated in the plot, colour bands representing the uncertainty [18].

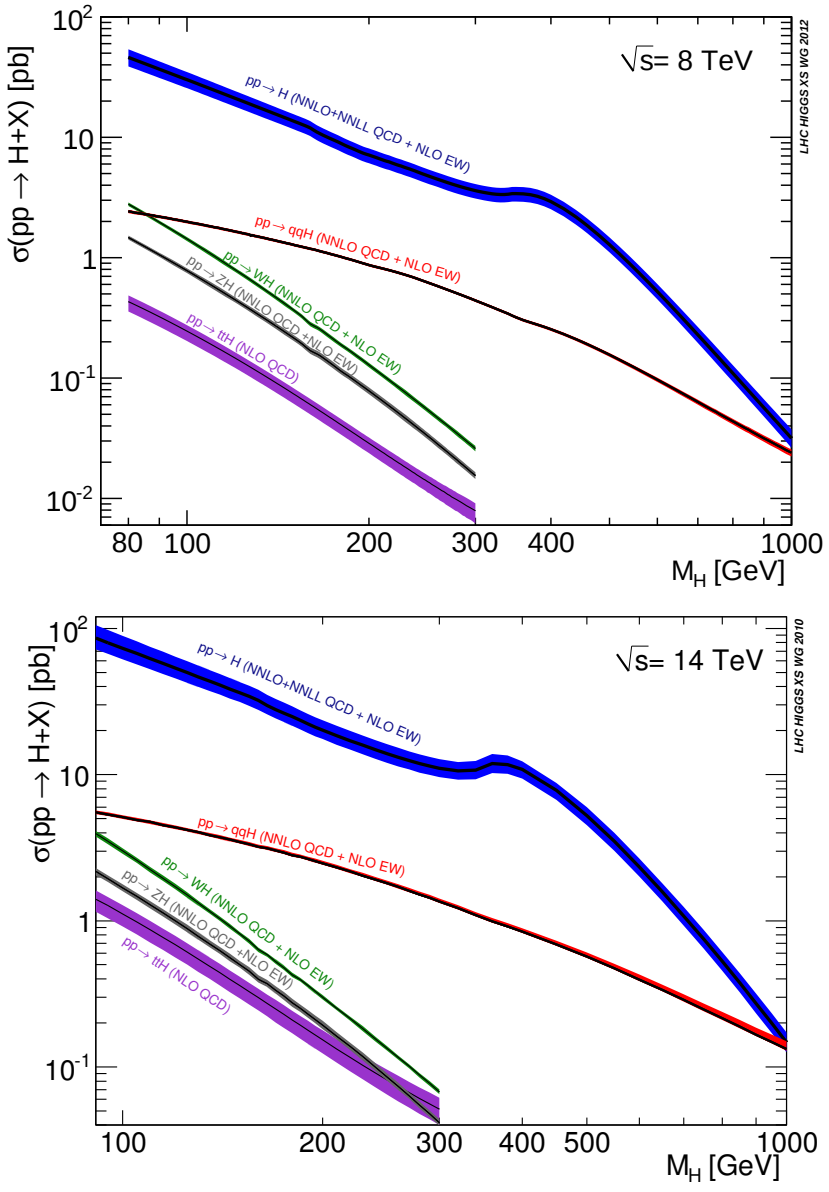
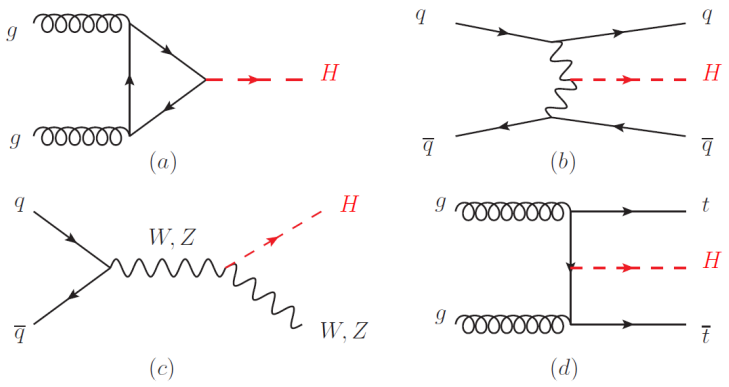


Figure 1.4 Feynman diagrams for processes contributing to the Higgs boson production in gluon fusion (a), vector boson fusion (b), Higgs Strahlung (c) and associated production with a top quark pair (d) [17].



Gluon Fusion The dominant process for Higgs boson production is the gluon fusion, with a cross section of almost 50 pb at $\sqrt{s} = 14$ TeV for a Higgs mass of $m_H = 125$ GeV. Due to the gluon being massless, there is no direct coupling between gluons and the Higgs boson. The process occurs through a quark loop and the dominant contribution arises from the top quark, because of its large mass. The Feynman diagram for Higgs production from gluon fusion is shown in Figure 1.4a. Despite its large production cross section, this production mode can be effectively studied only when the Higgs boson decays to particularly clean final states, allowing an efficient background rejection.

Vector Boson Fusion A Higgs boson can be produced from the fusion of two W or Z vector bosons radiated by initial state quarks. The Feynman diagrams for this production mode are drawn in Figure 1.4b. For a Higgs mass of $m_H = 125$ GeV, the cross section for Higgs production via vector boson fusion (VBF) is about 1.6 pb at a center of mass energy of 8 TeV, and increases to 4.2 pb at $\sqrt{s} = 14$ TeV. The two additional quarks in the final state are detected as two hard jets, with a tendency to be in the forward and backward regions of the detector. Thus, such particular topology allows for a sufficient background rejection, also depending on the Higgs decay mode.

Higgs Strahlung This process occurs when a virtual vector boson (V) decays to its on-shell state radiating a Higgs boson. Both the W and Z bosons contribute to this process, that is usually referred to as the VH production mode. The final state is therefore featured by the presence of either a W or a Z boson. Leptons from electroweak decays of vector bosons are particularly helpful for triggering and for background rejection in a hadronic environment, as they are easily identified and reconstructed. The leading order

Feynman diagram of VH production can be seen in Figure 1.4c. At a center of mass energy $\sqrt{s}$ = 14 TeV, the VH total cross section is $\approx$ 2.4 pb, of which about $\approx$ 0.9 pb due to the ZH process.

Associated Production with a Top Quark Pair Higgs radiation off top quarks, the latter produced mostly through gluon fusion, plays a role for light Higgs masses below $\approx$ 150 GeV at the LHC. A Feynman diagram for this process is shown in Figure 1.4d. The measurement of the $t\bar{t}$ H production rate can provide relevant information on the top-Higgs Yukawa coupling. For $\sqrt{s}$ =8 TeV, the $t\bar{t}$ H cross section is rather small compared to other Higgs production modes. However, it benefits from a relatively large increase when rising the center of mass energy. When passing from $\sqrt{s}$ =8 TeV to 14 TeV, the $t\bar{t}$ H production cross section increases from $\sim$0.13 pb to $\sim$0.6 pb. Therefore, also this is a promising channel for future Higgs physics analyses at the LHC.

Higgs Boson Decay Modes

The Higgs boson couples with bosons and fermions depending on their masses: the partial decay widths into massive vector bosons are proportional to the HVV coupling

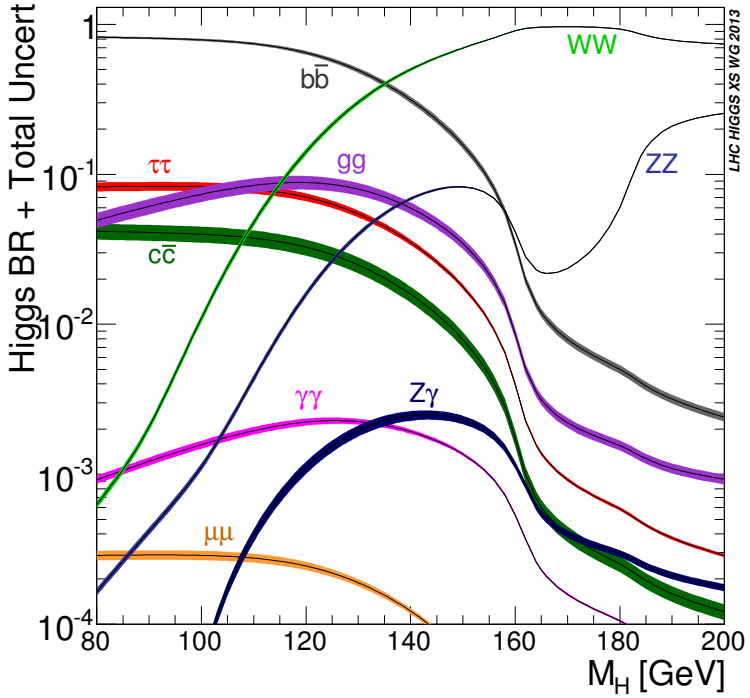
$$g_{HVV} = \left(\sqrt{2}G_F\right)^{1/2} m_V^2 \tag{1.27}$$

while the decay width into a pair of two same flavour fermions is proportional to the *Hff* coupling

$$g_{Hff} \propto \left(\sqrt{2}G_F\right)^{1/2} m_f \tag{1.28}$$

being m_V and m_f the masses of the vector bosons and fermions, respectively. Branching ratios (BR) for the different decay modes of the Higgs boson depends on its mass, as shown in Figure 1.5. The total decay width for a 125 GeV Higgs boson is $\Gamma_H = 4.07$ MeV, and the dominant decay mode is in a bottom quark pair, with a BR (H $\rightarrow$ $b\bar{b}$)$\approx$ 58%. The second most likely decay, with a BR of $\sim$20%, is the H $\rightarrow$ W W*, followed by the decays into gluons, taus, charm quarks and Z bosons pairs [19]. Despite its low branching ratio, the Higgs decay into a photon pair H $\rightarrow \gamma\gamma$ mostly through a top quark loop is one of the cleanest decay channels at the LHC. Together with the H$\rightarrow$ ZZ* decay channel in the four lepton final state, it provided the first significant evidences for a new resonance compatible with the Higgs boson.

Figure 1.5 Branching ratios for Standard Model Higgs boson decays, as a function of its mass [18].



1.5 The Higgs Boson Discovery at the Large Hadron Collider

Previously to the LHC start, most recent searches for the existence of the Higgs boson have been performed mainly at the Large Electron Positron (LEP) and the TEVATRON colliders.

The LEP was a e^+e^- collider operated at CERN between 1989 and 2000, reaching a maximum center of mass energy $\sqrt{s}$ = 209 GeV. The main purpouse of the LEP experiments was to perform precision measurement in the electroweak sector. Searches for the Higgs boson has been performed mainly in the e^+e^- →ZH production channel, that resulted [20] in a lower limit of 114 GeV for the Higgs mass, at a 95% confidence level (CL).

The TEVATRON, a proton-antiproton collider located at Fermilab near Chicago, operated at an energy up to $\sqrt{s} = 1.96$ TeV until 2011. There, the CDF and D0 experiments recorded data corresponding to an integrated luminosity of about 10^{-1} fb each. Direct searches for the Higgs boson have been performed mainly in the H→b$\bar{\text{b}}$, H→WW* and H → $\gamma\gamma$ decay channels. The existence of a Higgs boson with mass $m_H < 103$ GeV and $147 < m_H < 180$ GeV has been excluded at 95% CL. An excess of events in the region $110 < m_H < 145$ GeV, corresponding to a local significance of $\approx 3\sigma$ for $m_H = 120$ GeV, has been observed in the H→b$\bar{\text{b}}$ decay channel, in associated production with a vector boson [21][22].

In July, 2012 after about 3 years since the start of the LHC, both the ATLAS [7] and CMS [8] collaborations independently announced the discovery of a new resonance compatible with the Higgs boson. Such result was achieved combining searches in five decay modes: $\gamma\gamma$, ZZ*, WW*, $\tau\tau$ and b$\bar{\text{b}}$. For a given mass hypothesis, the sensitivity of a search channel depends on the production cross section of the Higgs bosons, its decay branching fraction, reconstructed mass resolution, selection efficiency and the background rejection. The most sensitive channels contributing to the discovery were the H$\rightarrow \gamma\gamma$ and H $\rightarrow$ZZ$^* \rightarrow 4l$ and will be discussed in the following.

Search for the Higgs Boson in the H$\rightarrow \gamma\gamma$ Channel

Despite having a branching ratio smaller than 0.3%, the H($\gamma\gamma$) channel is one of the most sensitive for the search of a light Higgs boson ($m_H < 150$ GeV), as the energy of the two photons can be directly measured in the electromagnetic calorimeter, with a relative resolution of 1-2% on their invariant mass and relatively small background. The main irreducible background for this channel arises from photon pairs produced in QCD processes and from γ+jets or dijets final states, when jets are misidentified as photons. In the diphoton invariant mass spectrum, the H($\gamma\gamma$) signal looks like a narrow peak over a smoothly falling background. Based on several kinematic variables, events are classified into different categories in order to separate contributions from the VH and VBF production modes. Remaining events are further categorized according to their expected $m_{\gamma\gamma}$ resolution and signal-to-background ratio. Categories with good mass resolution and larger signal-to-background ratio contribute most to the sensitivity of the search.

The diphoton invariant mass spectra measured from the ATLAS [23] and CMS [24] experiments are shown in Figures 1.6 and 1.7, together with the corresponding local p-values, that indicate the probability for a given excess of events to be compatible with a statistical fluctuation of the background. ATLAS observes an excess of events over background corresponding to local significance of 5.2 σ compared with 4.6 σ expected for a SM Higgs boson of mass $m_H = 125.4$ GeV. CMS observes its largest excess at $m_H = 124.7$ GeV with a significance of 5.7 σ compared with 5.2 σ expected for SM Higgs boson of that mass. The signal strength $\mu = (\sigma \times BR)/(\sigma \times BR)_{SM}$ which is the observed product of the Higgs boson production cross section (σ) and its branching ratio in units of the corresponding SM values, is 1.17 ± 0.17 for ATLAS and $1.14^{+0.26}_{-0.23}$ for CMS at the Higgs mass values mentioned above.

Search for the Higgs Boson in the H$\rightarrow$ZZ$^* \rightarrow 4l$ Channel

The search for the Higgs boson decay into two Z bosons, further decaying into four leptons, benefits from a high invariant mass resolution and a good background rejection. Three different final states $4e, 4\mu, 2e2\mu$ has been investigated, looking for a narrow resonance peak over a small continuous background mainly arising from non resonant

Figure 1.6 Left: invariant mass distribution of diphoton candidates obtained by the ATLAS experiment for the combined 7 and 8 TeV data samples. A fit to the data of the sum of a signal component fixed to m_H =125.4 GeV and a background component is superimposed. The bottom inset displays the residuals of the data with respect to the fitted background component. Right: local expected and observed significance for several Higgs mass hypotheses [23].

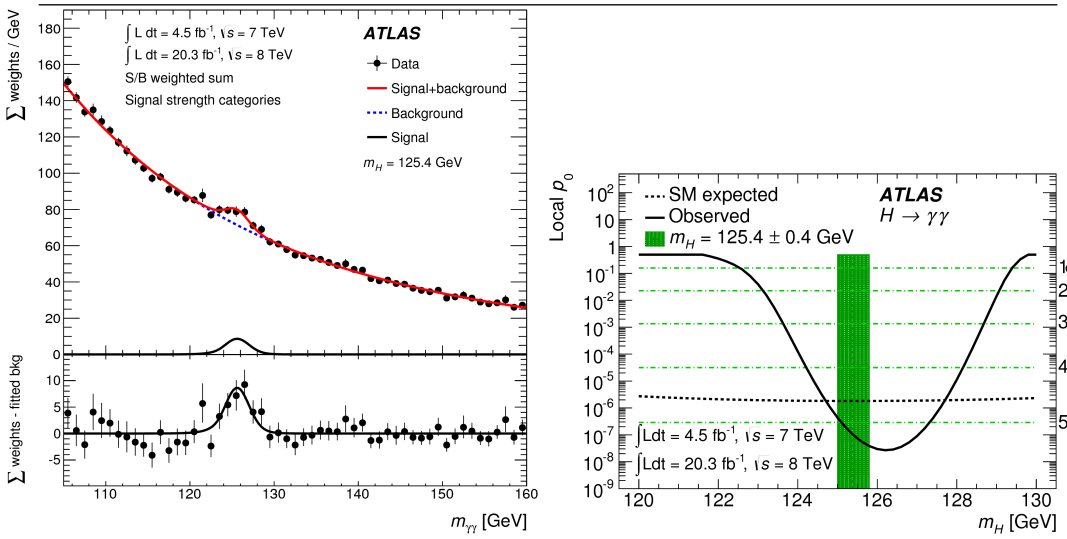
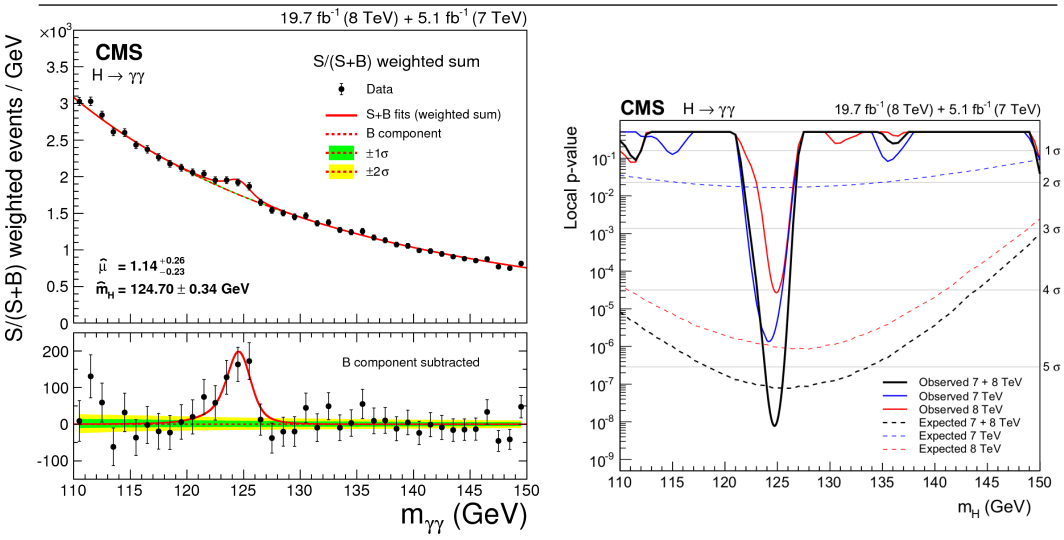


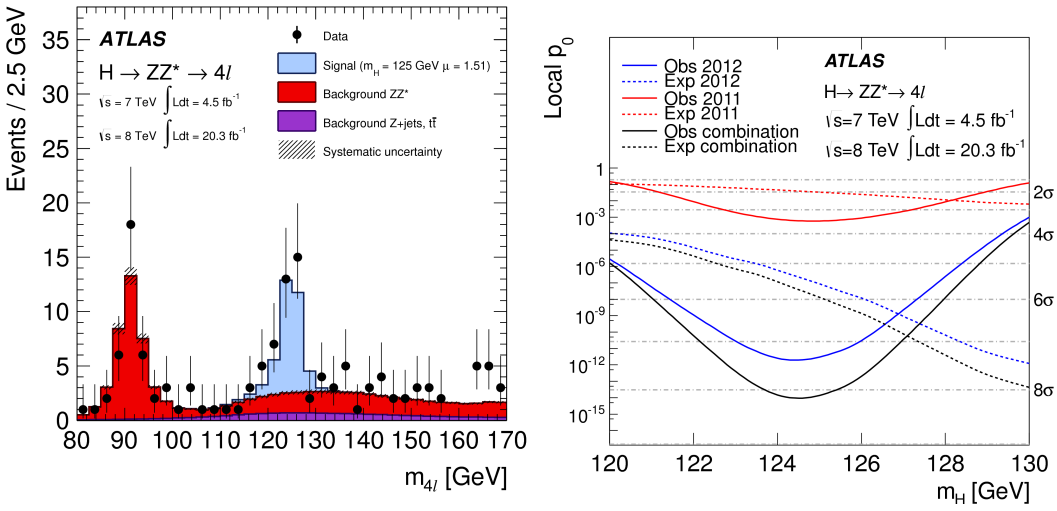
Figure 1.7 Left: CMS measurement of the diphoton mass spectrum weighted by the ratio S/(S + B) in each event class, together with the background subtracted weighted mass spectrum, for the combined 7 and 8 TeV data samples. Right: local observed and expected p-value for several Higgs mass hypothesis [24].



ZZ* production from $q\bar{q}$ annihilation and gluon fusion. Requirements on lepton isolation and impact parameter allow to reduce backgrounds from $t\bar{t}$ and Z+jets events. The kinematic of the four leptons is used for further discrimination between signal and background.

The four leptons invariant mass distributions as measured by the ATLAS and CMS experiments are shown in Figures 1.8 and 1.9 respectively. For the ATLAS collaboration, the largest deviation from the SM background-only expectation is observed at $m_H = 124.3$ GeV where the significance of the observed peak is 8.1 σ in the full 7 and 8 TeV datasets [25]. The expected significance for the SM Higgs boson at that mass is 6.8 σ. The CMS experiment observes its largest excess at $m_H = 125.6$ GeV with an observed significance of 6.8 σ to be compared with an expected significance of 7.2 σ at that mass.

Figure 1.8 Left: four lepton invariant mass distribution measured by the ATLAS experiment for the selected candidates compared to the background expectation for the 80-170 GeV mass range for the $\sqrt{s} = 8$ TeV and $\sqrt{s} = 7$ TeV combined data sets. The signal expectation for m_H =125 GeV is also shown. Right: local expected and observed significance for different Higgs mass hypotheses [25].



The best estimate for the Higgs boson mass obtained combining results from the ATLAS and CMS experiments, with the full 7 and 8 TeV datasets [27] is found to be $m_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.})$ GeV, as shown in Figure 1.10.

Figure 1.9 Distribution of the four lepton mass in the full mass range $70 < m_{4l} < 800$ GeV for the sum of the 4e, 2e2μ and 4μ channels, measured by the CMS experiment. The signal expectation is shown for a mass hypothesis of m_H =126 GeV. Right: significance of the local excess with respect to the SM background expectation as a function of the Higgs boson mass in the low mass region, for different likelihood fits. Reference results are obtained from the 3D fit $\mathcal{L}^{\mu}_{3D}$ [26].

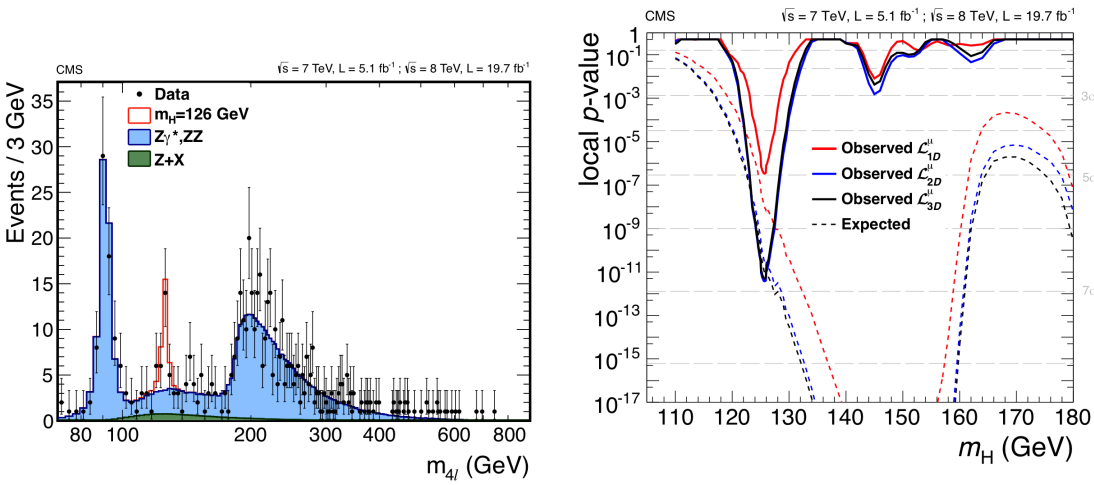
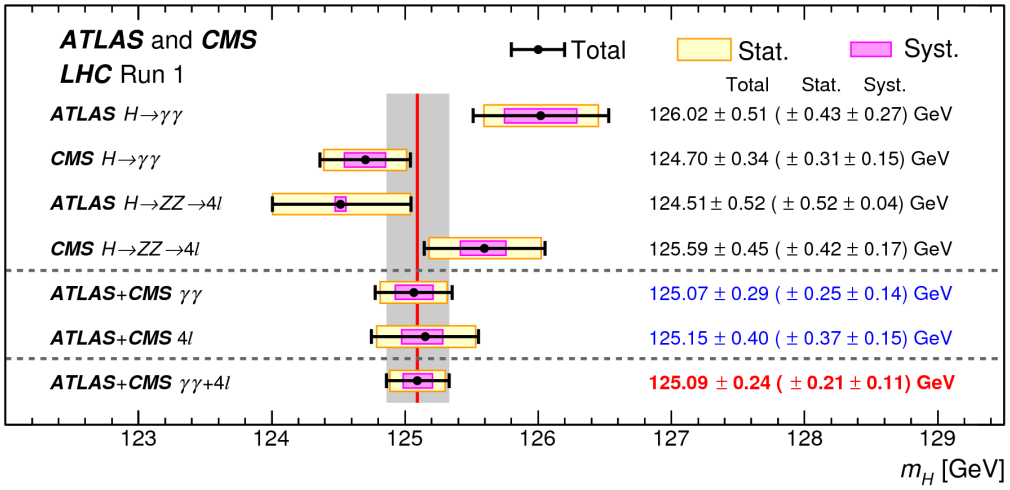


Figure 1.10 Summary of the CMS and ATLAS mass measurements in the $\gamma\gamma$ and ZZ channels, including the combined measurement. The systematic (magenta), statistical (yellow), and total (black) uncertainties are indicated. The (red) vertical line and corresponding (gray) shaded column indicate the central value and the total uncertainty of the combined measurement, respectively [27].



Chapter 2

The CMS Detector at the Large Hadron Collider

This Thesis work have been performed within the CMS Collaboration. The CMS experiment is located at the CERN Large Hadron Collider. In this Chapter, the main features of the LHC accelerator are briefly summarized. Then, the subsystems composing the CMS detector are described. Finally, the upgrades of the detector planned for the coming years are presented.

2.1 The CERN Large Hadron Collider

The Large Hadron Collider (LHC) is the most powerful particle accelerator built so far. It is located at CERN near Geneva in Switzerland and installed in the 26.7 km long tunnel which formerly hosted the Large Electron Positron collider (LEP). The LHC has been built for fundamental research in high energy physics, especially aiming at the discovery of the Higgs boson and at the search for new physics beyond the Standard Model.

LHC Design Parameters and Machine Layout

The LHC machine [28] is a two ring accelerator designed to collide proton bunches at a center of mass energy $\sqrt{s} = 14$ TeV and with an instantaneous peak luminosity $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1}. This latter parameter is defined as the proportionality factor between the rate of events produced in the collisions for a given process, and its cross section at the machine operating energy:

$$R = \mathcal{L}\ \sigma \tag{2.1}$$

The luminosity exclusively depends on the beam parameters; for identical bunches with a Gaussian profile and colliding head-on, it can be written as:

$$\mathcal{L} = f \frac{N^2 n_b}{4\pi\sigma_x\sigma_y} \tag{2.2}$$

where f is the revolution frequency, N the number of protons in each bunch, n_b the number of bunches per beam, σ_x and σ_y the transverse dimensions of the bunch.

At the LHC, the nominal revolution frequency is 11.3 kHz, with collisions occurring every 25 ns. The beam can be filled with a maximum of 2808 bunches, each containing nominally 1.15×10^{11} protons and with average transverse dimensions of $\sigma_x = 16$ μm and $\sigma_y = 100$ μm at the interaction points. The maximum particles density per bunch is limited by the nonlinear beam beam interaction experienced by the particles in the interaction region. It results in a tune shift that must be kept reasonably small in order to avoid beam losses due to resonances. If there is a non null crossing angle at the interaction point, the luminosity is reduced. At the LHC, the beam crossing angle is about 300 μrad, therefore a 15% decrease in the luminosity calculated according to Equation 2.2 must be taken into account [28].

The force needed to bend the 7 TeV proton beams along the desired trajectories is obtained with 1232 superconducting dipole magnets, each of them being 14.3 m long. They are cooled down to 1.9 K and produce a magnetic field of 8.3 T.

The basic layout of the LHC lattice is shown in Figure 2.1. It consists of eight arcs spaced by eight 500 meters long straight sections, that serves as experimental or utility insertion. The two beams circulates in two separated pipe lines, hosted in the dipole magnets, as can be seen in Figure 2.2. Figure 2.3 shows the two beam pipes and the field lines of the LHC dipole magnets. When approaching each of the four collision points, the beams are squeezed by dedicated quadrupole magnets in order to increase luminosity and are then collided.

Currently, four main experiments are located on the LHC, placed around the interaction points (IP). Figure 2.1 shows their location around the machine ring. The ATLAS [32] and CMS [33] detectors are located at IP 1 and 5 respectively. They are multipurpose detectors designed to record data in high luminosity regime. The LHCb experiment [34] focuses on B-physics and is located at IP 8. Moreover, during special runs the LHC is also operated with lead ion beams and the dedicated experiment called ALICE [35] is located at IP 2. In addition to the main experiments, other three experiments are intended for more dedicated studies: TOTEM [36] and LHCf [37] focuse on forward production, while MoEDAL [38] aims at searching for magnetic monopoles.

Figure 2.1 Schematic layout of the LHC. Beam 1 circulates clockwise and Beam 2 counterclockwise. The regions where the beams are squeezed and collided are indicated with a star sign. The positions of the four experiments on the interaction points are indicated [29]

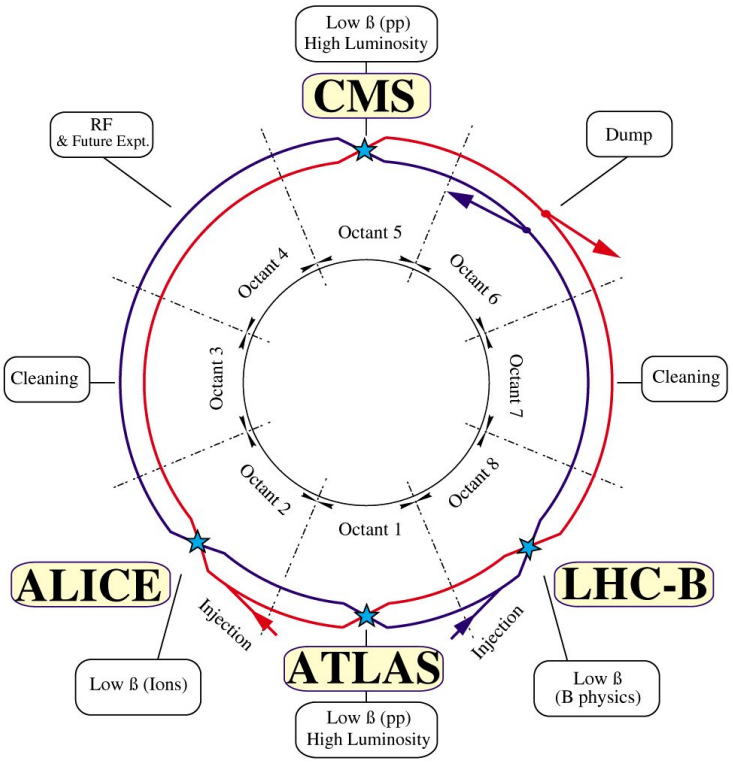


Figure 2.2 A LHC dipole cross section [30].

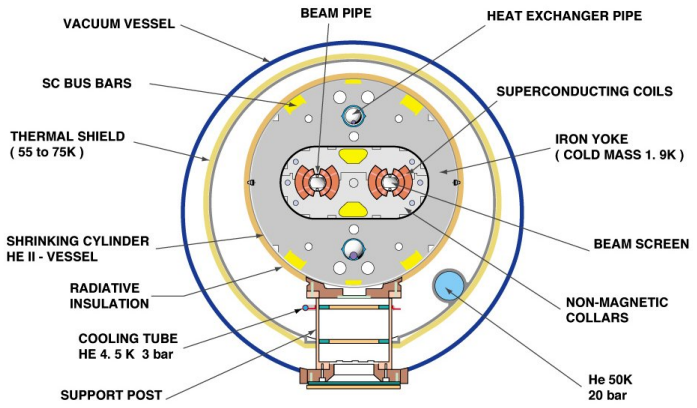
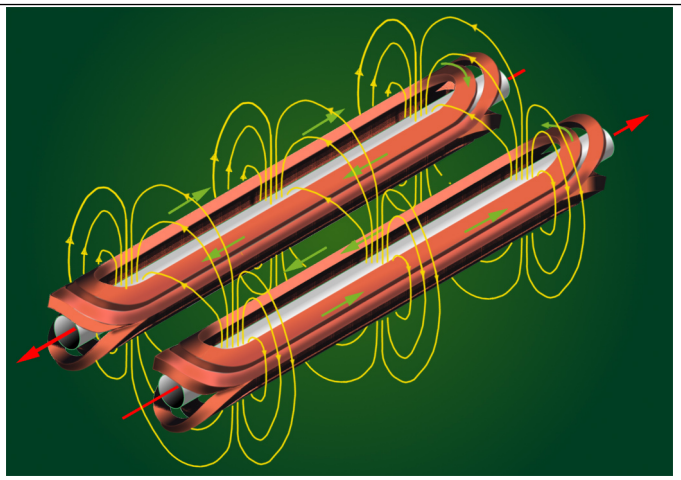


Figure 2.3 Beam pipes and magnetic field lines of the LHC dipole magnets [31].



LHC Machine Operation

Before being injected in the LHC, proton bunches are accelerated in many steps, passing from the lower to the higher energy machines of the CERN accelerator complex. Protons are obtained by a source of hydrogen gas, stripping hydrogen atoms of their electrons using an electric field. They are then accelerated in the Linac 2 linear accelerator, that brings their energy up to 50 MeV. The beam is then injected into the Proton Synchrotron Booster (PSB), which accelerates the protons to 1.4 GeV, followed by the Proton Synchrotron (PS), which pushes the beam to 25 GeV. Protons are then sent to the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV. Then, they are finally transferred to the two beam pipes of the LHC and accelerated to the desired energy thanks to radio-frequency cavities. Figure 2.4 summarizes the LHC injection chain. The minimum time required for ramping up the beam energy from 450 GeV to 7 TeV is approximately 20 minutes. Under normal operating conditions, beams circulate for many hours inside the LHC beam pipes.

The operation of the LHC [39] started in March 2010 with the first collisions of 3.5 TeV proton bunches, for a total center of mass energy of 7 TeV. An increase in beam energy to 4 TeV marked the start of operations in 2012 and the decision was made to stay at a 50 ns bunch spacing with around 1380 bunches. One of the main features of operation in 2011 and 2012 was the high bunch intensity (up to 150% of nominal with 50 ns bunch spacing) and lower-than-nominal emittances offered by the excellent performance of the injector chain of Booster, Proton Synchrotron and Super Proton Synchrotron.

The first phase of LHC operation, usually referred to as Run 1, lasted until December 2012. The total integrated proton-proton luminosity delivered to both the ATLAS and CMS experiments reached nearly 30 fb^{-1} and enabled the discovery of a Higgs boson. Figure 2.5 shows the integrated luminosity delivered to the CMS experiment in Run 1.

The LHC schedule foresees regular periods of shutdowns during which both the machine and the experiments detectors can be maintained and eventually upgraded. The first

Figure 2.4 The LHC injection chain [28].

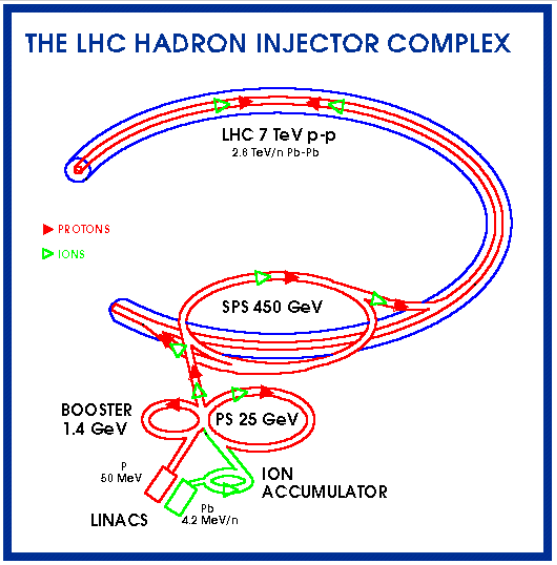
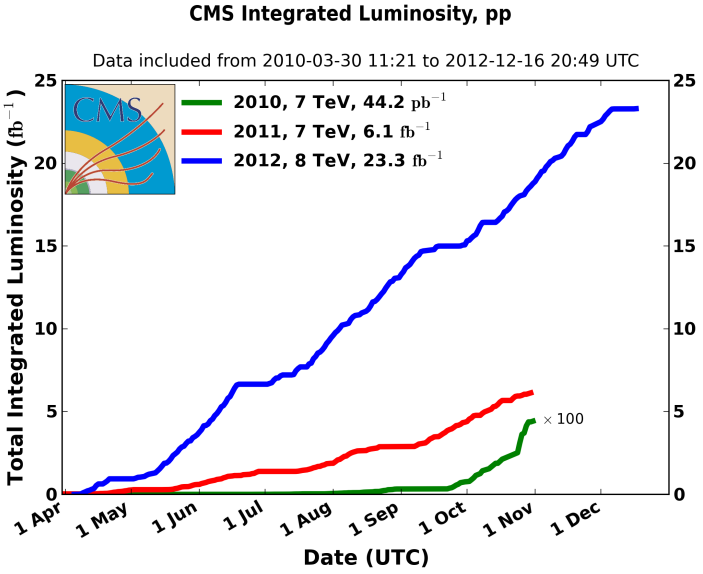


Figure 2.5 Integrated luminosity delivered to CMS during stable beams and for proton collisions, in data taking periods in 2010 (green), 2011 (red) and 2012 (blue) [40].



long shutdown (LS1) started in February 2013 and lasted about two years. First collisions at an energy of 13 TeV in preparation for the second physics run, also referred to as Run 2, started in May, 2015.

2.2 The Compact Muon Solenoid Detector

The Compact Muon Solenoid [33] is a general purpouse detector placed at IP 5 of the LHC. It has been designed to address a wide physics program such as tests of the Standard Model at the TeV scale and searching for the Higgs boson. In addition to that, extensive searches for hints of new physics beyond the Standard Model are being carried out, comprising searches for new particles predicted by supersymmetry and exotic models. Besides, precision measurements in the B-physics sector are of primary importance, as some rare decays might be enhanced due to the contributions of processes involving new particles.

In order to meet the above mentioned goals at the LHC environment, the detector must provide good particle identification and momentum resolution for muons and other charged particles, with high efficiency and over a wide geometrical range. Moreover, good electromagnetic and hadronic energy resolution and almost hermetic coverage are required. Due to the large flux of particles coming from the interaction region, both the detectors and the front end electronics must work in a highly irradiated environment.

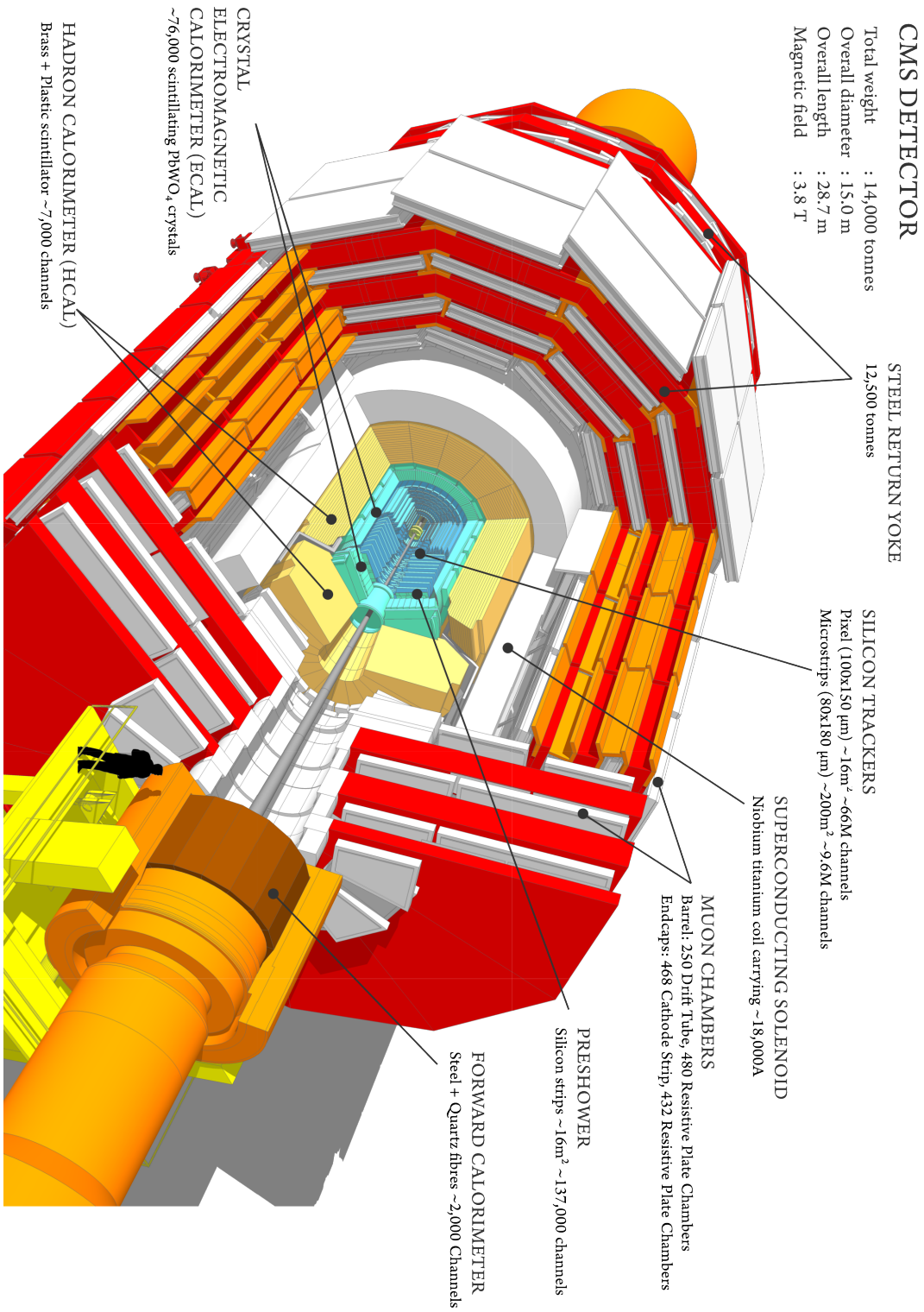
The adopted detector structure is cylindrically symmetric around the beam direction, with a big barrel region covering the central part and two endcaps that closes the structure on both sides.

The main distinguishing feature of the CMS detector is its compact structure that can be achieved thanks to the high solenoidal magnetic field. It provides large bending power needed to precisely measure the momentum of charged particles and allows muon detectors to be embedded in a iron yoke where the return field is confined.

In addition to that, CMS benefits from a full silicon-based inner tracker and a homogeneous electromagnetic calorimeter. The inner tracker and both the electromagnetic and hadronic calorimeters are placed inside the bore of the magnet coil, while the muon system is located outside of the magnet and is subject to the return magnetic field. The overall dimensions of the CMS detector are a length of 21.6 m, a diameter of 14.6 m and a total weight of 12500 tons. A schematic view of the whole detector showing the different components is given in Figure 2.6.

The following paragraphs provide an overview of the CMS magnet and sub-detectors, whose details can be found in Reference [33].

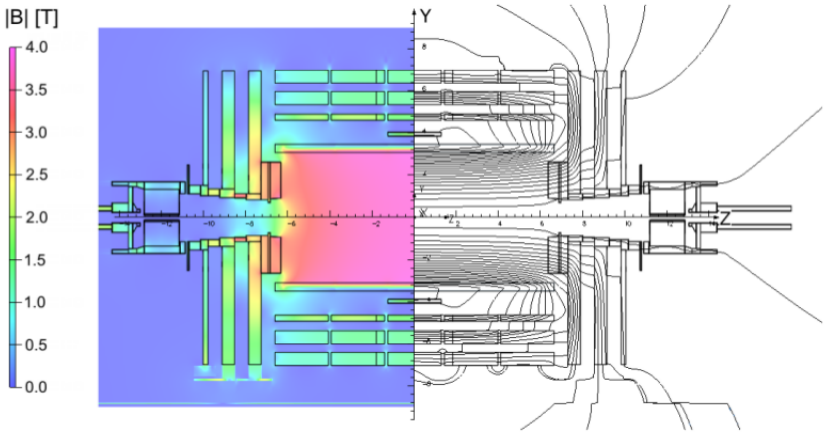
Figure 2.6 Sectional view of the CMS detector with all its subsystems. The LHC beams travel in opposite directions along the central axis of the CMS cylinder colliding in the middle of the detector [41].



2.2.1 Solenoid

The CMS magnet is a superconducting solenoid providing a magnetic field of 3.8 T. Magnetic field strength and lines on a longitudinal section of the CMS detector are shown in Figure 2.7. The coil has a inner radius of 5.9 m and a length of 12.9 m. It is cooled down to 4.5 K by a helium refrigeration plant and the whole structure is kept in insulation by two pumping stations providing vacuum on the cryostat volume. The field is closed by a 10000 tons iron return yoke made by five barrel wheels and two endcaps of three layers each. Muon stations are embedded between the yoke layers, exploiting the return field and enabling a very compact layout.

Figure 2.7 The magnetic field strength (left) and lines (right) of the CMS superconducting solenoid. Muon stations are embedded between the yoke layers, where the return field is exploited to bend muons [42].



2.2.2 Electromagnetic Calorimeter

A calorimeter is a detector meant to measure the energy of a particle. Electromagnetic calorimetry is based on the the production of a shower initiated by a photon or an electron inside the scintillating material. An efficient collection of the emitted light can be exploited to measure the initial particle energy. The Electromagnetic Calorimeter (ECAL) of the CMS detector is a hermetic homogeneous calorimeter composed by 61200 lead tungstate ($PbWO_4$) scintillating crystals in the central barrel region and 7324 crystals in each of the two endcaps. This material has been chosen because of its short radiation length (X_0=0.89 cm) and Moliere radius (2.2 cm), together with fast light emission and radiation hardness. The light yield of about 30 photons/MeV is relatively low and is collected by silicon avalanche photodiodes APDs (vacuum phototriodes in the endcaps). Both the crystals and photodiodes response are sensitive to temperature changes, requiring high temperature stability.

The barrel section has an inner radius of 129 cm and made up by crystals with a front cross section of about 22×22 mm^2, comparable to the typical transverse shower size in

$PbWO_4$, and a length of 230 mm, corresponding to $\sim 26\ X_0$. The barrel ECAL covers the pseudorapidity region up to $|\eta| = 1.479$.

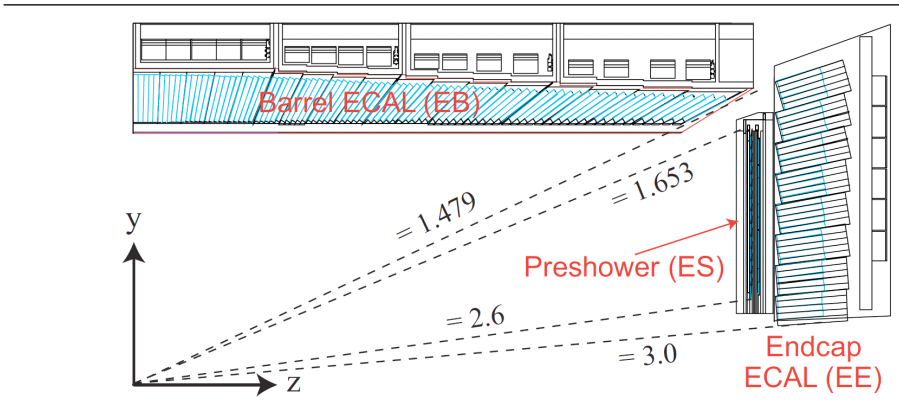
The endcaps, placed at a distance of 314 cm from the vertex and covering a pseudorapidity range of $1.479 < |\eta| < 3.0$, are each structured as two semi-circular aluminium plates containing structural units of 5×5 crystals, known as "supercrystals". They have a front cross section of $28.6 \times 28.6\ mm^2$ and a length of 220 mm ($24.7\ X_0$). A preshower detector designed for a more efficient π_0/γ separation is placed in front of the crystal calorimeter over much of the endcap η range. It is composed of two layers of lead absorber instrumented with orthogonal layers of silicon strip sensors. Figure 2.8 shows the arrangement of the ECAL crystals belonging to one quarter of the whole calorimeter.

The energy resolution of the CMS electromagnetic calorimeter can be decomposed in three different contributions:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{2.3}$$

where S and N are the stochastic and noise terms respectively, and C is a constant energy independent contribution. The values of these parameters have been obtained from Gaussian fit to the reconstructed energy distributions measured during tests with electron beams [43]. Results and dependence of the resolution as a function of energy are shown in Figure 2.9.

Figure 2.8 Geometric view of one quarter of the ECAL showing the arrangement of crystal modules in the barrel and endcaps region. The preshower detector is placed in front of the endcap ECAL [33].



2.2.3 Hadron Calorimeter

The CMS hadron calorimeter (HCAL) is a sampling calorimeter for hadron energy measurements. It must provide good containment and hermeticity, so that the measured $E_T{}^{miss}$ is actually due to non interacting particles (e.g. neutrinos), and not to a mismeasured hadronic energy. The HCAL surrounds the ECAL and is designed to maximize

the amount of material (quantified in interaction lengths) inside the magnet coil. An additional layer of scintillators, the hadron outer (HO) detector, is placed on the outer side of the solenoid.

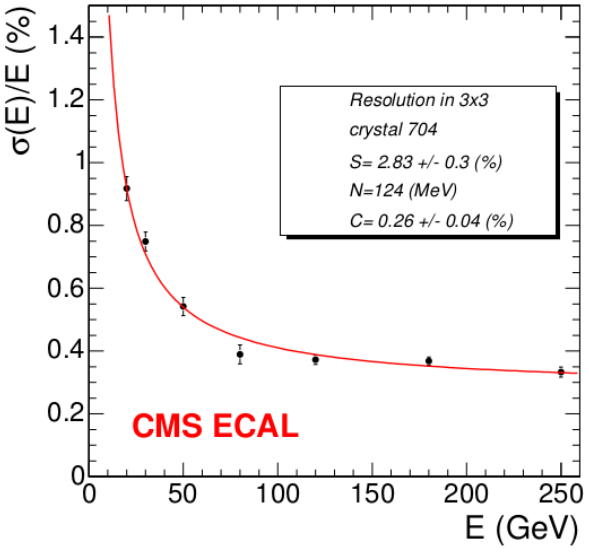
A sampling calorimeter has a layered structure, where layers of scintillating active material are spaced out by layers of absorber material. For the CMS HCAL, brass has been chosen as absorber because it has a reasonably short interaction length and is a non-magnetic material. The active material part consists of plastic scintillator tiles read out by wavelength-shifting fibers.

The hadron calorimetry system is organized in an inner hadron barrel, an outer detector, two endcaps and two forward calorimeters. The hadron barrel (HB) part consists of wedges covering the pseudorapidity region $|\eta| < 1.4$, the final granularity is $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$. The energy resolution in the barrel region is approximately $\sigma/E \sim 90\%/\sqrt{E}$.

The hadron outer (HO) detector is placed outside the magnet coil, in the wheels of the return yoke. It contains 10 mm thick scintillators with the same previous angular segmentation and covers the region $|\eta| < 1.26$. The HO samples the energy from penetrating hadron showers leaking through the back part of the barrel. Thanks to the HO, the effective thickness of the HCAL in the low η region exceeds 10 interaction lengths, with the benefit of improving energy resolution.

Each of the two hadron endcaps consists of 14 η towers covering the pseudorapidity region $1.3 < |\eta| < 3.0$. The granularity is higher in the outermost part, corresponding to smaller η, with $\Delta\phi$ from 5 to 10 degrees and $\Delta\eta$ from 0.087 to 0.35 at the highest pseudorapidities.

Figure 2.9 ECAL energy resolution as a function of electron energy as measured in a test beam. Measured values of the three parameters in Equation 2.3 are listed in the box [43].

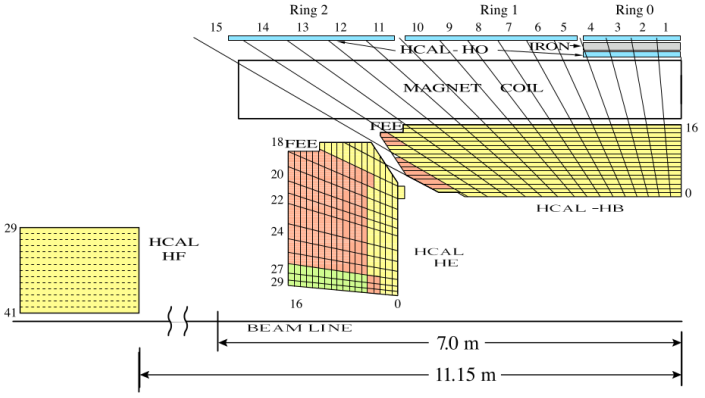


Finally, the hadron forward calorimeter (HF) ensures coverage up to $|\eta| = 5$. The forward region is characterized by a high flux of particles, therefore a steel/quartz fiber structure has been chosen. The absorber has a depth of 1.65 m, the diameter of the quartz fibers is 0.6 mm and they are arranged in a square grid, 5 mm apart from each other. The front face of the HF is 11.15 m far from the interaction point. The signal originates from Cherenkov light emitted in the quartz fibres, which is then channeled to photomultipliers.

The overall assembly enables the HCAL to be built with essentially no dead areas in ϕ, while there is a gap between the barrel and the endcap HCAL, through which the services of the ECAL and the inner tracker find their way out of the detector. The affected η region is $1.48 < |\eta| < 1.53$. A schematic view of the HCAL structure is shown in Figure 2.10.

The simulated and measured performance of the HCAL is shown in Figure 2.11 [44]. It shows the transverse energy (E_T) resolution of the hadronic shower (jet), as a function of the simulated transverse energy. It can be noticed how the granularity of the three parts of the HCAL has been chosen to result in a similar E_T resolution.

Figure 2.10 Schematic quarter view of the CMS hadronic calorimeter (HCAL) in the barrel (HB and HO), endcap (HE) and forward (HF) regions [44].



2.2.4 Muon Detectors

The systems for muon detection cover a large volume outside the magnet. The thickness of the inner detector, dominated by the HCAL, is such that almost all charged particles able to reach the muon detectors are actually muons. Three types of gaseous detectors are used to identify and measure them:

- drift tubes (DT) chambers in the barrel region ($|\eta| < 1.2$), where the expected particle rate is lower;
- cathode strip chambers (CSC) in the two endcaps (up to $|\eta| = 2.4$)

- resistive plate chambers (RPC) are used in addition to the above mentioned detectors, in both the barrel and endcap region up to $|\eta| < 2.1$, because of their fast response and good time resolution.

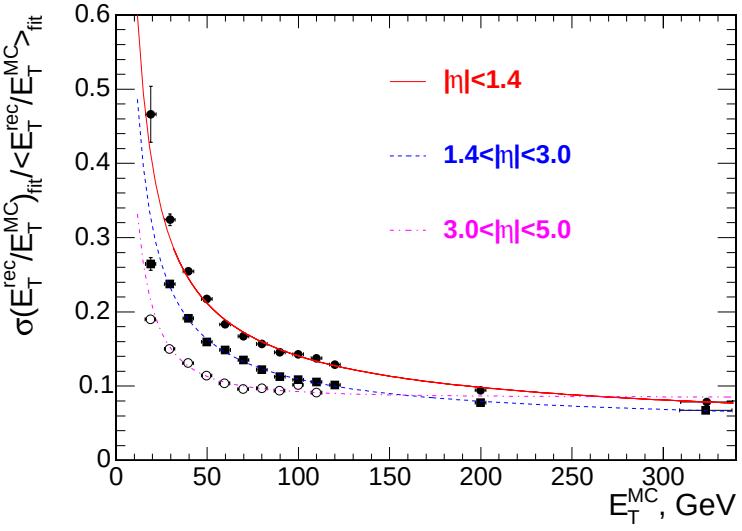
Combined information from RPCs and DTs (or CSCs) is used for the first level trigger system.

Figure 2.12 shows the different parts and arrangement of the muon system. In the barrel region, 250 drift tube chambers are grouped into 4 stations placed inside the magnet return yoke, extending radially from 4.0 to 7.0 m from the beam axis. Chambers in different stations are staggered not only to avoid dead spots but also to allow stand-alone bunch crossing tagging. The maximum drift length is 2.0 cm and the single-point resolution is approximately 200 μm. Each station is designed to measure the muon trajectory with a ϕ precision not worse than 100 μm in position and about 1 mrad in direction.

The Muon Endcap (ME) system comprises a total number of 468 CSCs. Every CSC has a trapezoidal shape and consists of six gas gaps, each gap having a plane of radial cathode strips and a plane of anode wires running almost perpendicularly to the strips. Charged particles traversing a plane of a chamber ionize the gas, and an the primary ionization charge is amplified close to the wire to produce an electron avalanche. The signal on the wires is fast and used for triggering, while the image charge induced on a group of cathode strips allows a precise position measurement by determining the centre-of-gravity of the charge distribution. Each CSC measures up to 6 space coordinates (r, ϕ, z). The spatial and angular ϕ resolution are about 200 μm and 10 mrad, respectively.

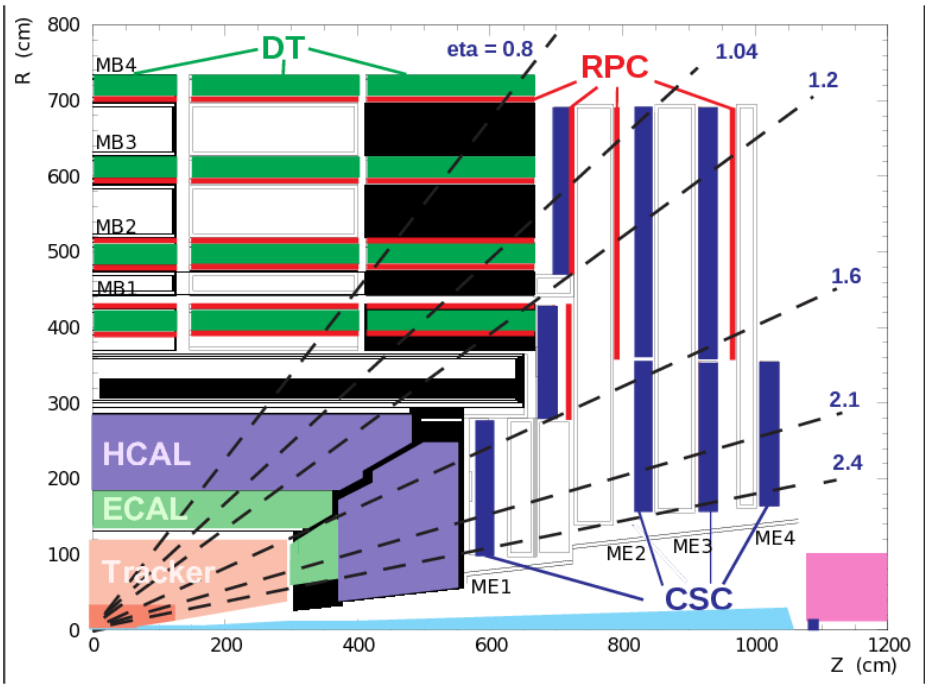
Muons produced in proton collisions with sufficient transverse momentum ($p_T \sim 2$ GeV) are detected both in the inner silicon tracker and in the outer muon detectors. The muon

Figure 2.11 The jet transverse energy resolution as a function of the simulated jet transverse energy for the three η regions [45].



transverse momentum can then be measured independently taking into account only the tracker or the muon chambers hits, or combining the information. The p_T resolution as a function of the muon p_T itself is shown in Figure 2.13, separately for the barrel and the endcap regions: for low p_T muons, the resolution is dominated by multiple scattering occurring in the traversed material, hence the tracker only measurement provides a better resolution. Muons with $p_T > 200$ GeV are much less sensitive to multiple scattering and benefit from the chamber high spatial resolution.

Figure 2.12 Longitudinal layout of one quadrant of the CMS detector. The four DT stations in the barrel (MB1-4, green), the four CSC stations in the endcap (ME1-4, blue), and the RPC stations (red) are shown [46].



The amount of material of the CMS detector in terms of interaction lengths after the ECAL, HCAL and each muon station is provided in Figure 2.14.

2.2.5 Tracker

The tracker is the innermost part of the CMS detector. It is a silicon detector with an active area of more than 200 m^2. The purpose of the tracker is to provide high precision three-dimensional measurements of the hits generated by charged particles. Reconstructed tracks are used to identify the primary and secondary vertices of the event and to measure particle momentum p_T. It can be obtained from the following Equation 2.4, once the radius of curvature R of the track is measured and the magnetic field strength B is known:

$$p_T\,[\text{GeV}] = 0.3 \cdot B\,[\text{T}] \cdot R\,[\text{m}] \tag{2.4}$$

Figure 2.13 Muon transverse momentum resolution, using either the muon system only, the inner tracker only or their combination, in the barrel (left) and endcap (right) regions [45].

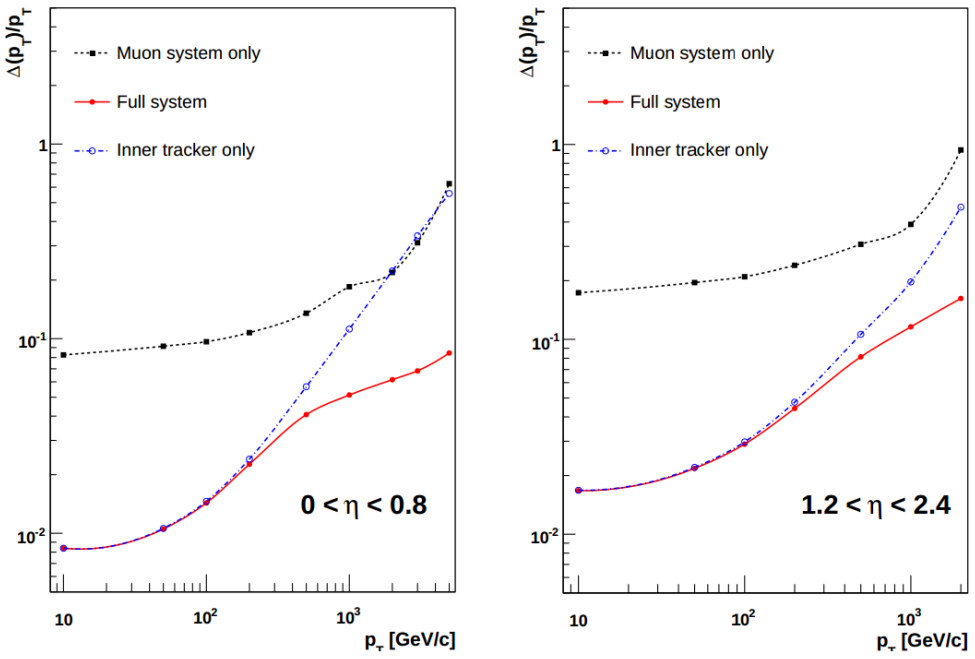
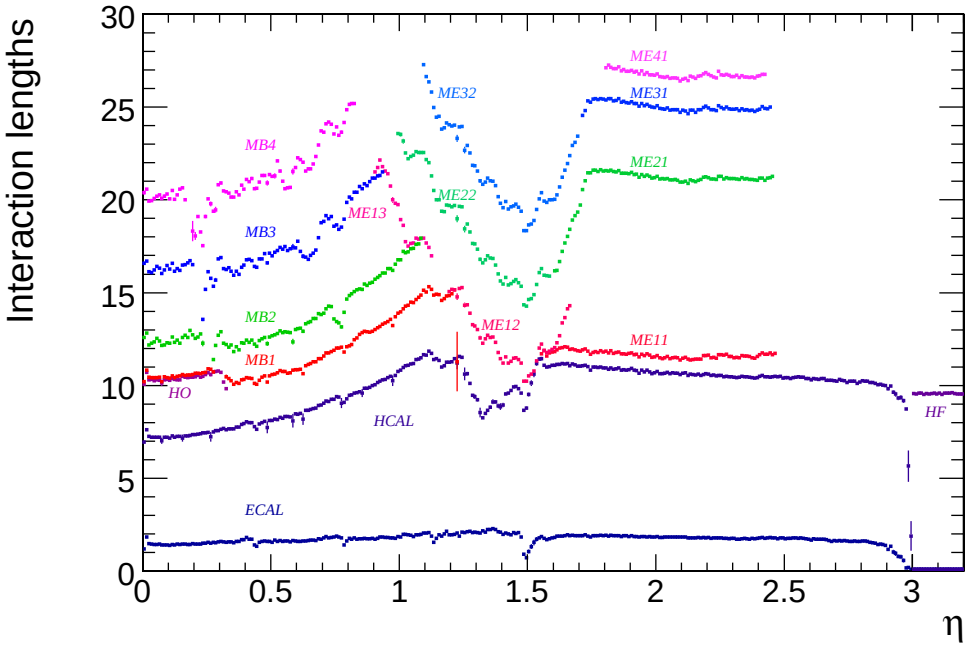
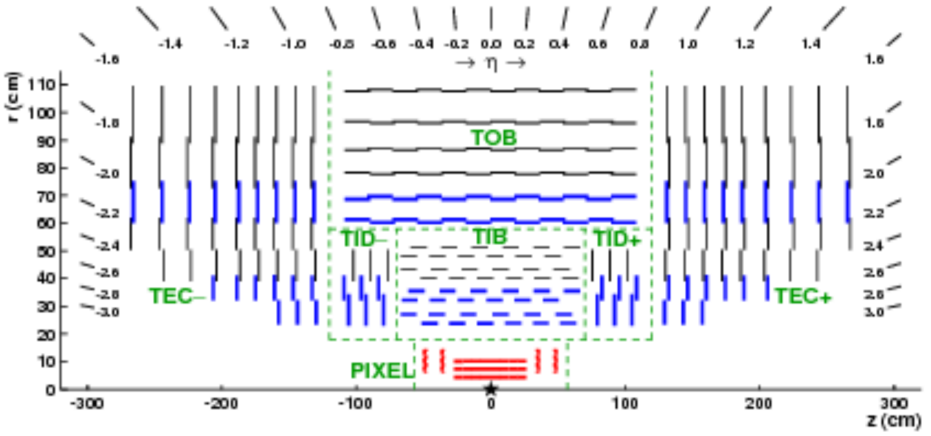


Figure 2.14 Material thickness in interaction lengths after the ECAL, HCAL, and at the depth of each muon station as a function of pseudorapidity [45].



The layout of the CMS tracker is shown in Figure 2.15. It extends radially from 4.4 cm to nearly 110 cm away from the interaction point, and its total length is approximately 540 cm. As other subsystems, it is also organized into a barrel and two endcaps regions. Both pixel and strip modules are used. In the innermost region, pixels are preferred because their smaller dimensions allow to keep low occupancy despite the high flux of particles and to achieve a better hit resolution. Silicon microstrips of two different sizes are used to cover the region starting from a radius of 20 cm.

Figure 2.15 Schematic view of the CMS tracker in the $r-z$ plane. The pixel detector (red) has three layers in the barrel and four disks in the endcap region. Strip tracker modules are in black and blue (stereo modules) [47].



Pixel Tracker

The pixel detector consists of three cylindrical barrel layers with two endcap disks on each side. Barrel layers are located at a mean distance of 4.4 cm, 7.3 cm and 10.2 cm from the interaction point, and have a length of 53 cm. The two endcap disks, extending from 6 to 15 cm in radius, are placed on each side at $|z| = 34.5$ cm and 46.5 cm.

In order to achieve the optimal vertex position resolution, a design with an almost square pixel shape of 100×150 μm^2 has been adopted. The barrel comprises 768 pixel modules arranged into half-ladders of 4 identical modules each. The endcap disks comprise 672 pixel modules arranged in blades, each of them composed by seven modules. The overall layout of the pixel detector is sketched in Figure 2.16.

The overall area covered by the pixel modules is about 1 m^2. Each pixel cell is bump bonded to a read out chip (ROC), there are approximately 16000 ROCs for a total of about 66 million channels. The coverage of the pixel tracker extends up to $|\eta| = 2.5$.

The spatial resolution is measured to be about 10 μm for the r-ϕ measurement and about 20 μm for the z measurement. Such resolutions are achieved not only thanks to the small pixel size but also exploiting the charge sharing among neighbouring pixels induced by the CMS magnetic field [48], in which the tracker is immersed.

Figure 2.16 View of the CMS pixel tracker, with the three barrel layers (green) and two encap disks on each side (pink) [45].

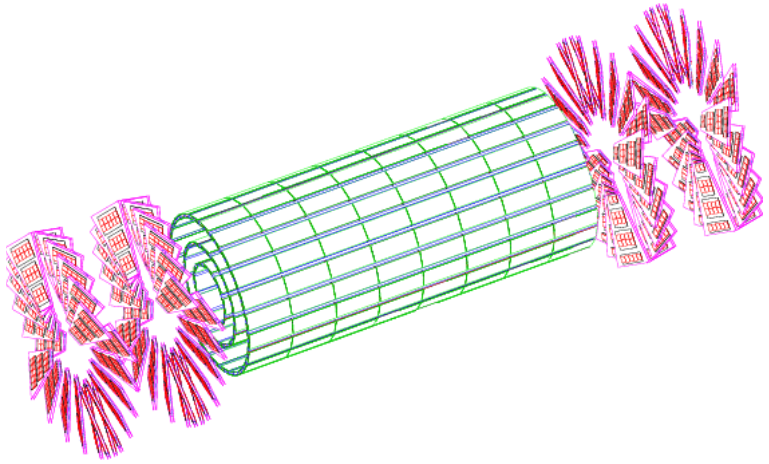
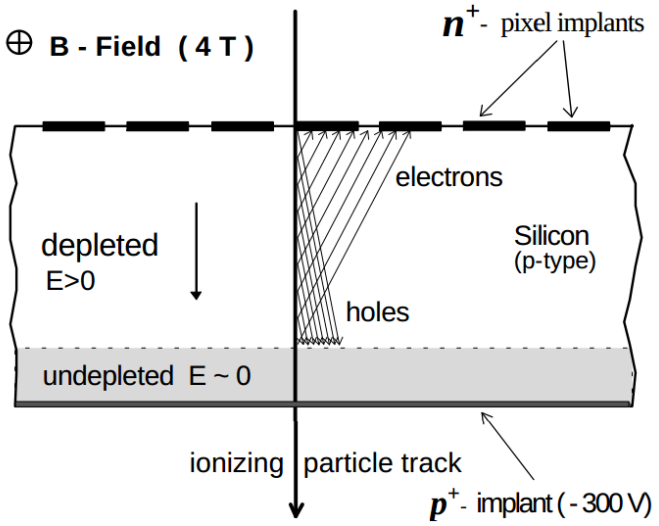


Figure 2.17 Charge sharing induced by deflection of charge carriers by the CMS magnetic field. Drift trajectories are not perpendicular to the detector surface but tilted of an angle called Lorentz angle [49].



As can be seen in Figure 2.17, charge carriers produced in the tracker volume do not drift along the electric field lines but are deflected by the Lorentz force. Their trajectories are tilted of an angle called Lorentz angle, resulting in a sharing of the collected charge among more close-by pixels. The hit position is obtained evaluating the centre of gravity of the collected charge over the pixels, therefore improving the resolution. The magnitude of the Lorentz angle depends on the bias electric field and on the magnetic field applied. For the CMS pixel tracker, it has been measured to be $\sim 25°$ in the barrel region [50]. In order to have charge sharing also in the endcap disks, the forward detectors are tilted of 20° in a turbine-like geometry.

Strip Tracker

The entire silicon strip detector consists of almost 15400 modules, mounted on carbon-fiber structures and housed inside a temperature controlled outer support tube. Silicon microstrip modules are placed from a radial distance of 20 cm up to 110 cm from the interaction point. The tracker barrel part is separated into an Inner (TIB) and an Outer Barrel (TOB). In order to avoid excessively shallow track crossing angles, the Inner Barrel is shorter than the Outer Barrel, and there are three additional inner disks on each side in the transition region between the barrel and the endcaps.

The TIB is made of four layers and covers up to $|z| < 65$ cm, using silicon sensors with a strip pitch which varies from 80 to 120 μm. The first two layers are made with *stereo* modules, consisting of two silicon sensors mounted back-to-back with their strips aligned at a relative angle in order to provide a measurement in both $r-\phi$ and $r-z$ coordinates. An angle of 100 mrad has been chosen, resulting in a single-point resolution of between 23-34 μm in the $r-\phi$ direction and 230 μm in z. The TOB comprises six layers with a half-length of $|z| < 110$ cm. The strip pitch varies from 120 to 180 μm. Also for the TOB the first two layers are rotated by the same stereo angle as for the TIB to achieve a single-point $(r-\phi, z)$ resolution of 35-52 μm and 520 μm respectively.

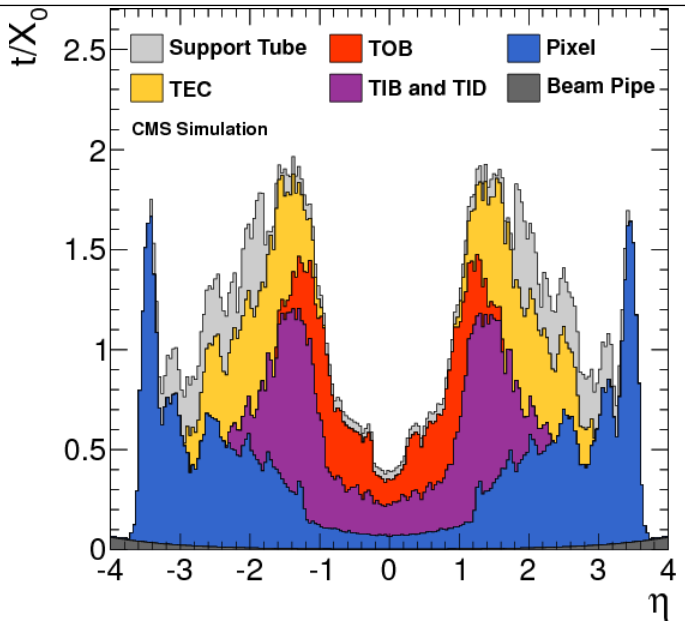
The endcaps are divided into the TEC (Tracker End Cap) and TID (Tracker Inner Disks). Each TEC comprises 9 disks that extend into the region $120 < |z| < 280$ cm, and each TID comprises 3 small disks that fill the gap between the TIB and the TEC. Modules are arranged in rings and have strips that point towards the beam line, therefore having a variable pitch.

The material budget of the CMS tracker in units of radiation length X_0 as a function of η is provided in Figure 2.18, showing the contributions from the different tracker sub-detectors.

Tracker Operation in Run 1

The reliability of the detector was constant over the data taking period despite the challenge presented by the continuously increasing instantaneous luminosity and absorbed dose. By the time of the shutdown in 2013, about 2.3% of the barrel and 7.2% of the endcap modules of the pixel detector were inactive, mostly due to faulty wire-bonds or poor connections. Over the same period of data-taking, about 2.5% of the strip detector became inactive due to short-circuits in the control rings and HV lines, or as a result of faulty optical communications. Repair of the damaged modules was part of the maintenance performed during 2013; up to 1.5% of the pixel barrel, 0.5% of the pixel endcap and 1% of the strip detectors have been recovered [52].

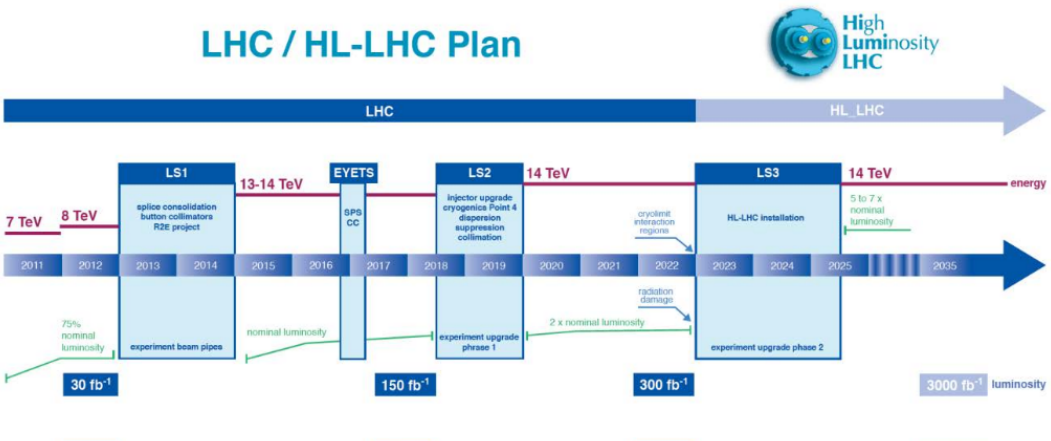
Figure 2.18 Material budget of the CMS tracker in units of radiation length X_0, as a function of η [51].



2.3 Upgrades of the CMS Detector

Figure 2.19 shows the LHC timeline from 2011 to 2035, indicating the center of mass energy and peak luminosity foreseen in different operation phases and the technical stop periods planned for upgrade and consolidation of the LHC machine as well as of the experiments.

Figure 2.19 Timeline of the LHC machine, from the beginning of its operation in 2011 to the end of the High Luminosity Phase in 2035. Technical stops and long shutdowns (LS) are needed to upgrade of the LHC machine and of the experiments [53].



After the first long shutdown (LS1) in 2013/2014, the LHC machine has been restarted in spring 2015 and will run for almost two years with gradually higher energy (from 13 to 14 TeV) and instantaneous luminosity up to at least $\sim 10^{34}$ cm^{-2}s^{-1}. In such conditions,

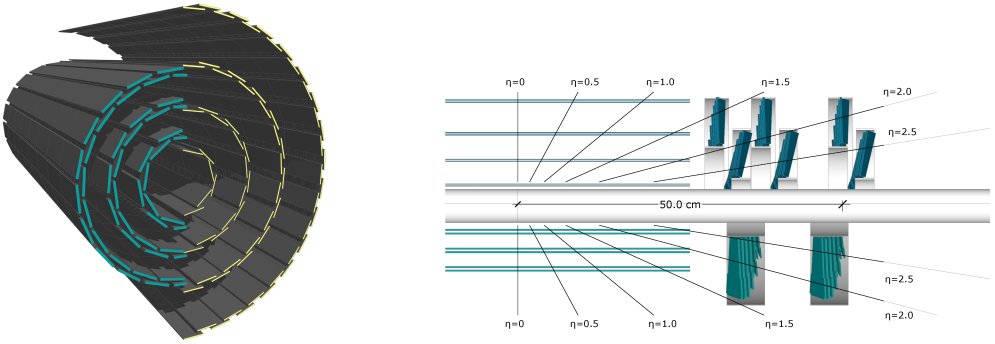
a maximum pile up of 50 interactions is foreseen at the end of 2016, before the end of the year technical stop. The third long shutdown (LS3) scheduled for 2023 is meant to significantly upgrade the LHC machine and prepare it for a period of high luminosity (HL) run starting around 2025, with an instantaneous luminosity of $\leq 5 \times 10^{34}$ cm^{-2}s^{-1} and average expected pile up of 140 interactions.

The large increase in luminosity requires upgrades of the experiments, in order to retain good performances in the more challenging environment. The CMS collaboration foresees to improve the detector in several stages during the shutdowns. The Phase 1 upgrade [54] is the first upgrade program taking place in the end of the year technical stop of 2016 and in the second long shutdown in 2018/2019. It will be followed by the Phase 2 upgrade during the third long shutdown in 2023/2025.

2.3.1 Phase 1 Upgrade

During the shutdowns foreseen in the Phase 1 operation, the electronics for the trigger system, the hadron calorimeter and muon detectors will be updated. The muon system will be equipped with a fourth layer of cathode strip chambers and, in the endcap region, of resistive plate chambers [54].

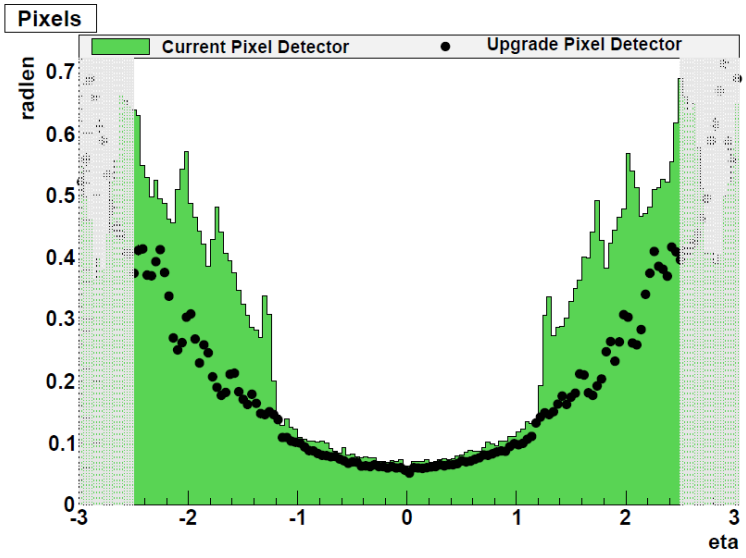
Figure 2.20 The pixel detector geometry before and after the Phase 1 upgrade, in the barrel (left) and in the endcap regions. Left: the current barrel geometry (blue) is place aside the upgraded four-layer geometry (yellow). Right: the present tracker design (bottom) compared to the upgraded geometry (top) [55].



However, the most important change will affect the pixel tracker, that will be replaced in the technical stop at the end of 2016. The current inner tracker was designed to sustain a fluence of 6×10^{14} hadrons per cm^2, corresponding to the first four years of LHC operation, and to operate at a maximum luminosity of 1×10^{34} cm^{-2}s^{-1}, that will be likely exceeded at the end of LS2. When keeping the current pixel system, readout inefficiencies would decrease the ability of reconstructing tracks, and radiation damages would spoil the hit resolution. In fact, higher bias voltage is needed to keep good charge collection efficiency in irradiated sensors, but the increase of the electric field will result in a smaller Lorentz angle, leading to a reduced charge sharing among neighbouring pixels. Therefore, the pixel tracker needs to be replaced.

The design of the Phase 1 pixel detector compared to the current pixel tracker is shown in Figure 2.20.

Figure 2.21 Material budget for the Phase 1 pixel detector (black dots) as a function of η, measured in radiation lengths, compared to the present pixel tracker (green) [55].

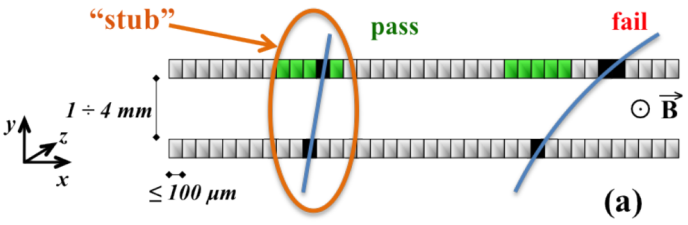


It will consist of four barrel layers, instead of the present three, and of three endcap disks on each side. The innermost layer will be placed at a radial distance of 3.0 cm, instead of 4.4 cm of the current tracker, requiring a new beam pipe with a smaller diameter. The addition of the fourth layer will reduce the distance between the last pixel hit and the strip system, improving track extrapolation towards the silicon strip tracker. Moreover, the additional layers will give four hit coverage over the full $|\eta| < 2.5$ range making it easier to detect triplet track seeds used for track reconstruction, therefore reducing track inefficiency. The pixel size will remain unchanged at 100×150 μm^2, and the number of pixels to be read out will therefore increase from 48 to 79 million. Despite the addition of one layer in the barrel and in each endcap disk, the amount of material in the central η region will be reduced by building a lighter mechanical support, changing the cooling system (from C_6F_{14} to CO_2) and moving services further out in z from the interaction point. Figure 2.21 compares the material budget in the present and upgraded pixel detector in the full η coverage. A description of the module layout, comprising the silicon sensor the read out chip characteristics is presented in Chapter 3.

2.3.2 Phase 2 Upgrade

During the long shutdown scheduled for 2023, the CMS detector will be prepared to operate with an extremely high instantaneous luminosity, resulting in an expected pile up of 140 interactions. The most relevant improvements to the detector in the Phase 2 upgrade [56] include the replacement of the electromagnetic calorimeter in the endcaps, possibly covering a region up to $|\eta = 4$, and the installation of GEM detectors in the forward regions, to increase trigger redundancy and extend muon detection up to $|\eta| = 4$.

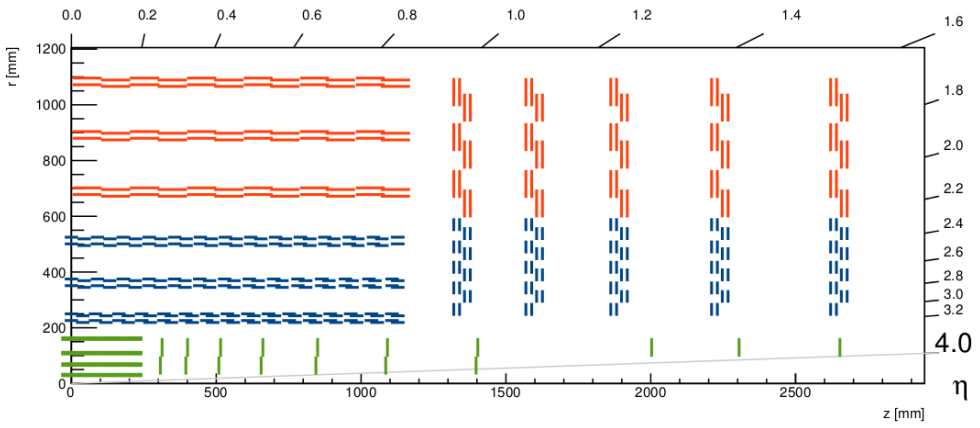
Figure 2.22 The concept of *s*tubs: correlation of signals in closely-spaced sensors enables rejection of low-p_T particles by defining an acceptance window for the width of the stub [56].



In addition to that, the entire silicon tracker will need to be replaced, as both the silicon strip tracker installed in 2008 and the Phase 1 pixel tracker will have reached the design lifetime in terms of radiation hardness. The pixel detector design is not yet decided but a very similar geometry to the one chosen for the Phase 1 upgrade is plausible. In order to cope with the increased particle density the pixel cell has to be reduced. Thanks to the advancements in readout electronics, it is now feasible to design the readout cell with a 65 nm technology, allowing for a cell size of 25×100 μm^2 or 50×50 μm^2 that would increase the radiation tolerance of the detector. For the innermost layer an alternative design using 3D coloumnar silicon sensors has been considered, offering higher radiation resistance.

The overall concept of the outer tracker is already well defined and somewhat different from the current one. The outer tracker will in fact play a key role in the Level 1 trigger by providing a first measurement of the track transverse momentum. This is planned to be achieved with the so called p_T modules, composed by two closely spaced silicon sensors, read out by a common front, aiming to correlate the hits in the two sensors and select pairs (stubs) compatible with particles above the chosen p_T threshold. Figure 2.22 shows the stubs generated from high and low-p_T particles. A low energetic particle will bend more, and the hit in the second sensor will be further displaced from the hit in the first sensor, with respect to a high momentum particle. Depending on the spacing between the sensors, on the local magnetic field an on the desired p_T threshold, an acceptance window for the stub width can be defined to reject low momentum tracks.

Figure 2.23 Layout of the Phase 2 tracker. The pixel detector is drawn in green, the outer tracker is in blue (PS modules) and red (2S modules). The tracker coverage extends up to $|\eta| = 4$ [56].



Two types of modules are being developed:

- 2S modules to be used in the outer layers, consisting of two superimposed strip sensors of approximately 10 $\times$ 10 cm^2. Strips of 5 cm length and 90 μm pitch are mounted parallel to one another
- PS modules, to be placed in the barrel layers located at a radial distance of about 20 to 60 cm from the interaction point, where the particle density is lower. They are composed of two sensors of $\sim$ 5 $\times$ 10 cm^2. One is segmented in strips and the other one in macro-pixels of 100 μm $\times$ 1.5 mm, providing sufficient granularity for distinguish close-by tracks and triggering in a relatively high track density condition. Moreover, the PS modules layout allows precise measurement of the z coordinate needed for primary vertex discrimination for the Level 1 trigger.

The electronics needed for stub reconstruction is hosted on each module. The increased extension of the tracker will cover the region up to $|\eta| = 4$. The entire tracker layout is sketched in Figure 2.23. The pixel detector is drawn in green, while the outer tracker, consisting in 12 barrel layers and 10 endcap disks on each side, is in blue (PS modules) and red (2S modules). The feasibility of a less conservative geometry with tilted barrel layers is also under study: it would reduce the number of modules without affecting the tracker coverage and reduce the effective amount of material on the tracks path, but is more challenging for the cooling system.

Chapter 3

Phase 1 Upgrade of the CMS Pixel Tracker

The upgraded pixel detector will have four cylindrical layers in the barrel region and three disks in each side, as detailed in the previous Chapter. In the barrel part (BPIX) one additional layer will be placed at large radius to reduce the present gap between the last pixel and the first strip tracker layer. Moreover, the first pixel layer will be placed 10 mm closer to the beam pipe. Pixel modules will be mounted on a low mass structure with integrated cooling tubes. Each layer is organized into two halves, that are joint together. The mechanical support is designed such that the innermost layer can be extracted and eventually replaced without disconnecting any other layer.

The efforts of production and quality test of pixel modules are shared among several centers, following the scheme reported in Table 3.1. Barrel layers are numbered from 1 to 4, the first layer being the innermost one. Layer 1 and 2 are being produced in Switzerland by the PSI and ETH institutes, one half of Layer 3 is being build from a consortium of INFN sections in Italy and the second half at CERN, by a collaboration among CERN, Taiwan and Finland. The outermost layer is being produced in Germany by the DESY group in Hamburg, together with other German universities. Endcap disks are produced in the United States, at Purdue and FNAL.

However, all sensors and read out chips used to build different layers of the pixel tracker share the same design. To guarantee a uniform high quality of the final assembled detector, a precise quality assurance procedure to be applied by every production center has been established. It consists of several tests to monitor the quality of the module throughout the whole production chain, from its basic parts to the different assembly steps of components with specific functionality.

This Thesis work included active participation in the BPIX upgrade program currently going on in the Pisa production laboratory, therefore this activity will be discussed in more detail in this Chapter.

Table 3.1 Overview of the barrel pixel modules production for Phase 1 upgrade, including the number of modules per layer and the location of production centers.

Production of Barrel Pixel modules for the Phase 1 Upgrade

Layer	Radius	N. of modules	Production center
1	39 mm	96	Switzerland
2	68 mm	224	Switzerland
3	109 mm	352	Italy (50%) - CERN, Taiwan and Helsinki (50%)
4	160 mm	512	Germany

3.1 Layout of Pixel Modules

The modules designed for the pixel barrel (BPIX) layers consist of several parts connected to each other [55]. Figure 3.1 represents an expanded view of a BPIX module, showing the basic components. Following the scheme of Figure 3.1, from bottom to top, they are:

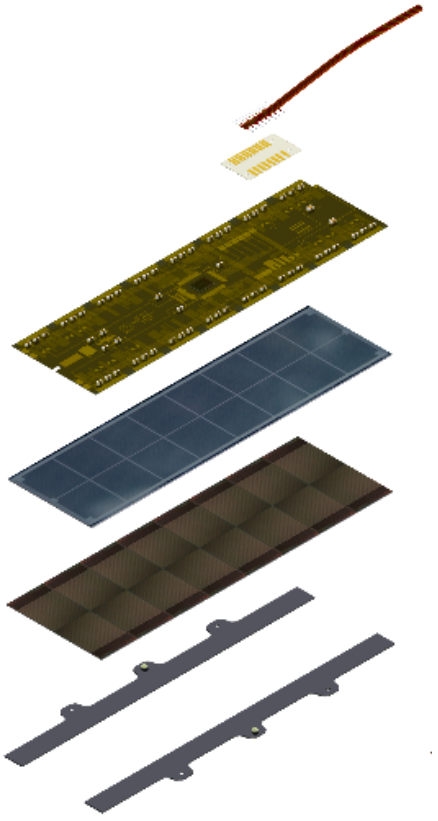
- base strips, two rods of Si_3N_4 to be glued on the back side of the ROC wafer, designed in order to allow mounting the modules on the mechanical structure
- the Read Out Chip (ROC) arranged in 2 rows of 8 ROCs, for a total of 16 per module
- the silicon sensor, organized into 16 matrices of 52 columns and 82 rows of pixels, each matrix connected to a different ROC. The size of the pixel cell is $100 \times 150\ \mu m^2$, unchanged compared to the current pixel sensor design
- the high density interconnect (HDI) circuit that provides signal and power to the chips will be glued on top of the sensor and wire bonded to the corresponding ROCs pads
- the Token Bit Manager (TBM) chip, glued and wire bonded to the HDI, controls the readout of ROCs and distributes clock and trigger signals
- the cable and connector that will distribute the power and allow the data acquisition for the module.

3.2 Silicon Pixel Sensor

The silicon sensor is the key component of the pixel module. The main advantages in using silicon are:

- the small charge collection time, thanks to the high mobility of charge carriers
- its high density, resulting in a high number of electron-hole pairs created by a ionizing particle traversing it ($\sim$ 396 eV/μm corresponding to 110 electron-hole pairs/μm for a minimum ionizing particle)
- the good intrinsic energy resolution provided by a low ionization energy ($\sim$3.6 eV)

Figure 3.1 Components of a barrel pixel module (Layer 2 to 4) for the Phase 1 upgrade. From bottom to top: base strips, ROCs, silicon sensor, HDI, TBM chip and the cable and connector. Layer 1 modules have the same structure but the base strips are placed along the short module sides and two TBMs on a modified HDI circuit are used. [55].



As the pixel detector is placed very close to the beam pipe, it has to operate in a harsh radiation environment. Figure 3.2 shows that the charged hadron fluence through the pixel tracker is expected to reach $\sim 10^{15}$ /cm^2 after a delivered integrated luminosity of $5 \cdot 10^5$ pb^{-1}, corresponding to 10 years of LHC activity at nominal design conditions.

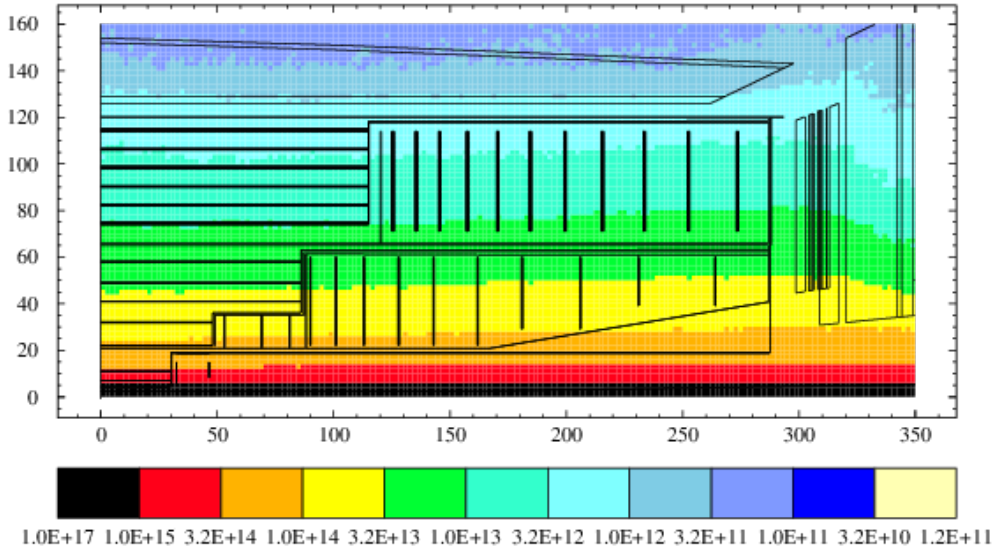
Particle irradiation has three main macroscopic effects in reversed bias silicon detectors:

1. the increase of the leakage current, leading to a rise of the sensor temperature due to higher power consumption. Cooling is needed to prevent a thermal runaway of the device.

2. introduction of deep energy levels in the band-gap with the creation of acceptor traps. This doping build up will imply a type inversion in the bulk substrate for n-type sensors that will become p-type

3. a decrease of the charge collection efficiency due to trapping of charge carriers by crystal defects. Such trapping is more effective for holes than for electrons.

Figure 3.2 Charged hadron fluences (per cm^2) in the tracker. Values are given for an integrated luminosity of $5{\cdot}10^5$ pb^{-1}, corresponding to 10 years of LHC operation at nominal design conditions [57].

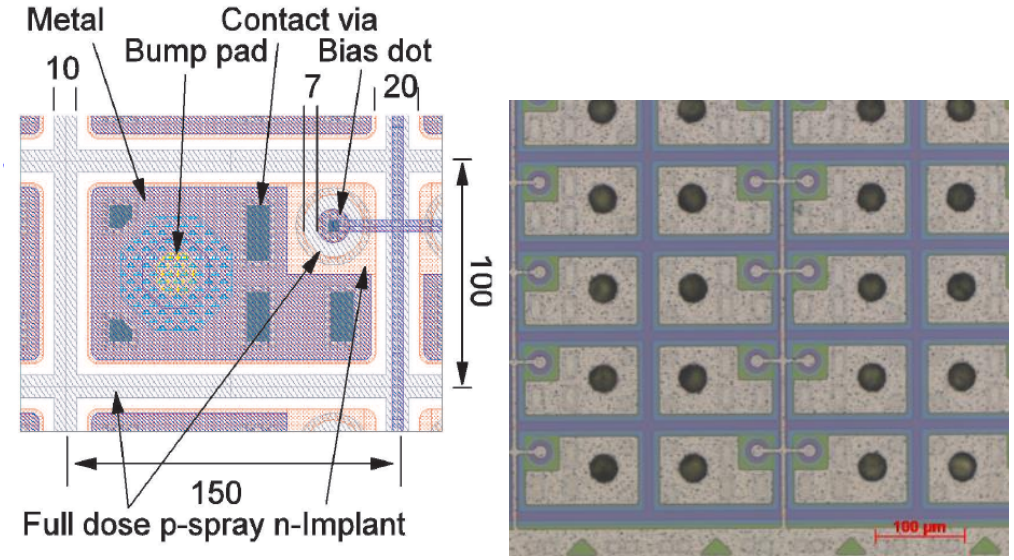


The last two items motivate the choice of n-in-n type silicon sensors adopted by CMS. In this configuration, the signal is formed by collecting electrons, that are less prone to trapping thanks to their higher mobility with respect to holes. In addition to that, the detector is still well performing after type inversion of the bulk from n to p-type. The n-in-n technology consists in having highly doped n implants (n^+), the pixel unit cell, on a n-type silicon bulk. A large p^+ implant on the back side provides the pn junction. The pixel side is covered by layers of insulating silicon oxide (SiO_2) and silicon nitride (Si_3N_2), protecting the segmented area, with openings designed to contact the n^+ pixel and extract the signal through metal electrodes layers.

Due to a static positive charge exposed by the silicon oxide at the interface with the bulk, a layer of electrons accumulates underneath the oxide and the pixel implants are no longer isolated from each other. To avoid that, the accumulation layer is interrupted

by p-impurities diffused around the pixel implant (p-spray). Moreover, the use of p-spray is advantageous for mitigating surface radiation damage consisting in ionization of the silicon oxide layer. Figure 3.3 shows the pixel layout on the n-side, with the detailed structure of a Pixel Unit Cell (PUC).

Figure 3.3 Left: layout of the pixel cell on the n-side, showing the bump pad, the bias dot and the location of p-spray. Dimensions are given in μm. Right: arrangement of pixel cells on the n-side. One sensor contains 16 ROCs, each of them having 52×80 pixel cells.



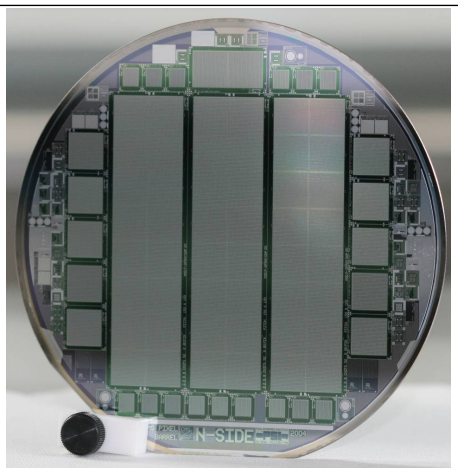
Pixel sensors are fabricated on 4 inches round wafers, with a thickness of 285±15 μm. They are obtained from a silicon monocrystal, whose elementary cell is cubic with coordination numbers <111>. Boron and phosphorus are used as p and n type dopants, respectively. The bulk is characterized by high resistivity $\sim$ 3.7 kΩcm and the voltage needed to fully deplete the junction is $\sim$ 60 V.

Each wafer accomodates three sensors, as can be seen in Figure 3.4. A sensor is divided into 8×2 arrays of 4160 pixels, formed by 52 rows and 80 columns. A module has therefore 66560 pixels and a total active area of 16.2×64.8 mm^2.

3.2.1 Tests of Pixel Sensors

The first sensor quality tests are performed already on wafer, right after production. They consist in a visual inspection of the pixelated n-side and in a reverse I-V curve measurement. All sensor wafers for the italian half of Layer 3 are tested in Pisa in a clean room laboratory, equipped with a Karl Suss PA150 probe station and diagnostic instrument setup.

Figure 3.4 A silicon wafer of the Phase 1 production, with three pixel sensors and service structures for the quality assurance process. The picture shows the n-side.



Optical Inspection

The optical inspection is meant for identifying defects on the pixel layout that might compromise operation:

- layer defects occurring in the processing steps, which could affect sensor functionality or spoil the quality of the connection between the affected pixel and its readout cell. An example of this kind of defect can be seen in the left picture of Figure 3.5.
- pattern irregularities occurring during the lithography process, e.g. due to a mask misalignment, eventually modifying the local electric field. Such defects have not been observed.
- scratches due to inappropriate handling, as the ones photographed in the right picture of Figure 3.5.

The n-side of sensor wafers is inspected with an optical microscope embedded in the probe station. The magnification used is $\times 50$, that is combined with an internal magnifying power of $\times 10$, resulting in a total magnification of $\times 500$.

The kind of observed defects and their location on the sensor are recorded, as they can cause read out errors of the affected pixels.

IV Measurement

The IV test consists in measuring the leakage current as a function of the applied reversed voltage. The measurement can be done on wafer, having three needles contacting the p-sides of each sensor and a fourth needle to contact the common side, as can be seen in the picture of Figure 3.6.

Figure 3.5 Example of defects on pixel sensors found during optical inspection on the n-side: a residual of resistive material partially covering the bump area (left) and a scratch over the bump pad (right).

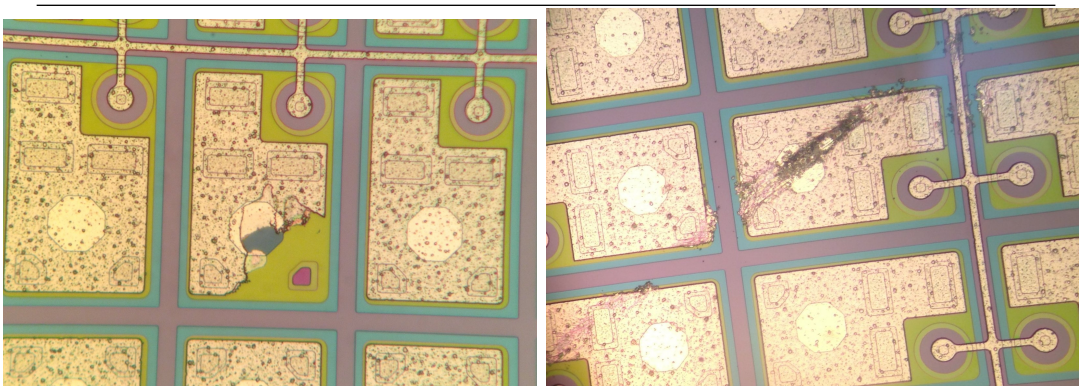
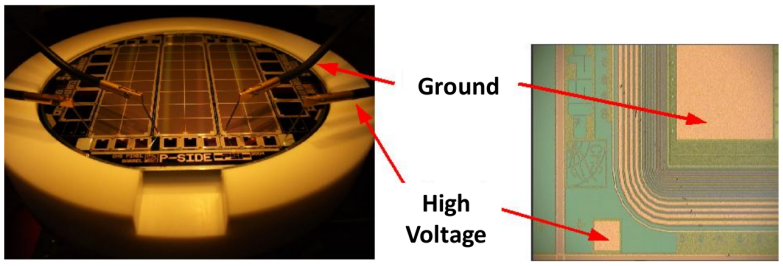


Figure 3.6 IV measurement on wafer. Left: the silicon wafer is placed on the vacuum chuck and contacted with four needles, one providing the high voltage bias and three connecting each sensor to ground. Right: wafer pads where the bias and ground needles are connected.



The voltage is increased in steps of 5 V, from 0 to 600 V, with a waiting time of 3 s between steps. The current measurement is repeated 10 times and averaged. Once the measurement is finished, the voltage is gradually ramped down. If the leakage current becomes too high ($> 100\,\mu$A), the measurement is automatically stopped in order to prevent high current and consequently energy dissipation through the junction. Basing on the IV test, sensors are accepted and can be used for production, or rejected. An accepted sensor must satisfy all the following criteria:

1. the leakage current at 150 V, $i(150) < 2\,\mu$A. The current has a dependence on the temperature, and this value refers to T = 20°C. Therefore, the effective temperature must be checked every time before performing the measurement and the measured currents scaled to the reference temperature of 20°C by using the following relation [48]:

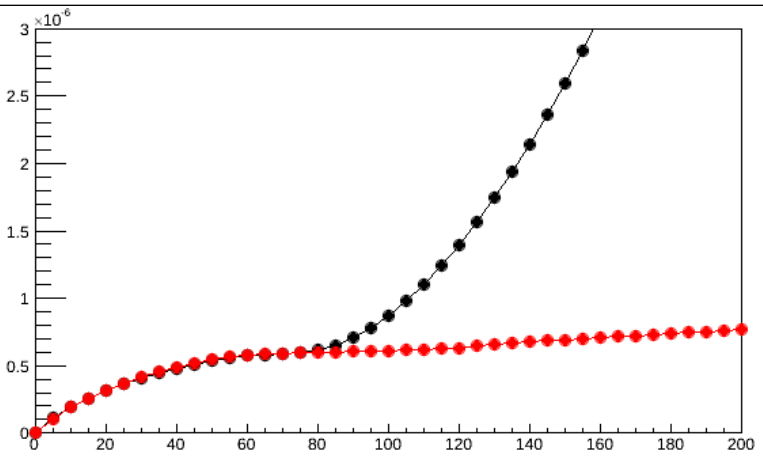
$$i(\mathrm{T}) \propto \mathrm{T}^2\, e^{-\frac{E_g}{2k_B \mathrm{T}}} \tag{3.1}$$

where $E_g = 1.12$ eV is the silicon energy gap and k_B is the Boltzmann constant.

2. the current increase must be far from a breakdown behaviour, with a relative increase $i(150\text{ V})/i(100\text{ V})$ smaller than 2. Being a ratio between two currents measured in the same conditions, it does not depend on the temperature.

An example of typical IV curves for a sensor passing or not the requirements are shown in Figure 3.7.

Figure 3.7 An example of IV curve for a silicon sensor on wafer passing the quality requirements (red) or not (black). The leakage current [A] is plotted as a function of the bias voltage [V].



The first IV measurements of sensors on wafer are performed by the vendor (CiS), up to a bias of 200 V. If at least two sensors match the above mentioned criteria, the wafer can be considered as good for detector production and is shipped to Pisa, where the IV curve is measured possibly up to 600 V. The two measurements show good agreement and the grading of the sensor remains unchanged for almost all (>98%) sensors. In other words, no unexpected behaviour appears for a good sensor when expanding the range of bias voltage up to 600 V. This demonstrates the effectiveness of the acceptance cuts, based on the production experience of the present vertex detector on a total of about 1400 sensors. Such acceptance cuts have been also tuned in order to maintain high voltage stability of pixel sensors after the foreseen radiation damage.

At the time of writing this Thesis, all sensors needed for production have been delivered by the vendor and qualified. Table 3.2 summarizes the quality assurance results for sensors that are being used by the Italian consortium for building one half of barrel Layer 3. Out of 318 sensors on wafer delivered by the vendor, only 49 of them did not meet the requirements and were rejected. The fraction of good sensors is ∼85%. Figure 3.8 shows the IV curves for 25 randomly chosen sensors, passing the quality requirements. The observed behaviour is very similar for all sensors, especially in the most important region for bias voltages below 200 V. Quantitatively, considering all 269 good sensors, the average current at 150 V bias voltage is found to be $6.13 \cdot 10^{-7}$ A, with a RMS of $2.34 \cdot 10^{-7}$ A, therefore confirming the uniformity of the production process.

Table 3.2 Quality assurance results of sensors on wafer, indicating the number and relative yield of good (rejected) sensors fulfilling the requirements (or not).

Quality Assurance results of sensors on wafer

	Good	Rejected	Total
	269	49	318
Fraction	84.6%	15.4%	

Figure 3.8 IV curves of 25 different sensors fulfilling the quality requirements. The leakage current [μA] is plotted as a function of the bias voltage [V], from 0 to 600 V (left). A zoom on the x axis up to 200 V (right) shows a similar behaviour for all sensors, indicating the uniformity of the production.

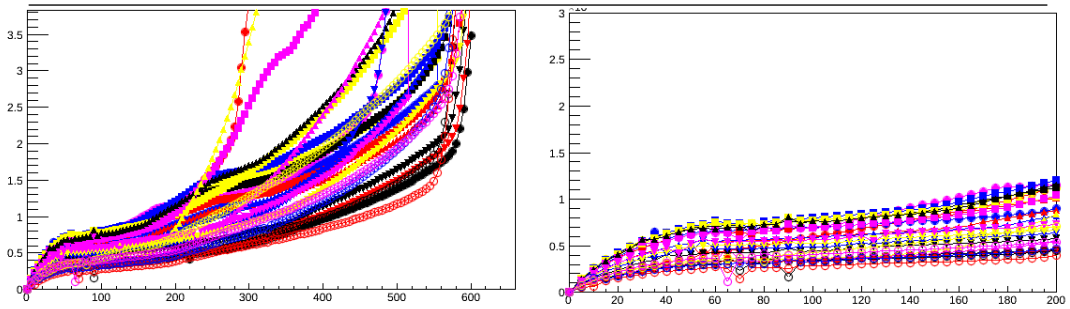
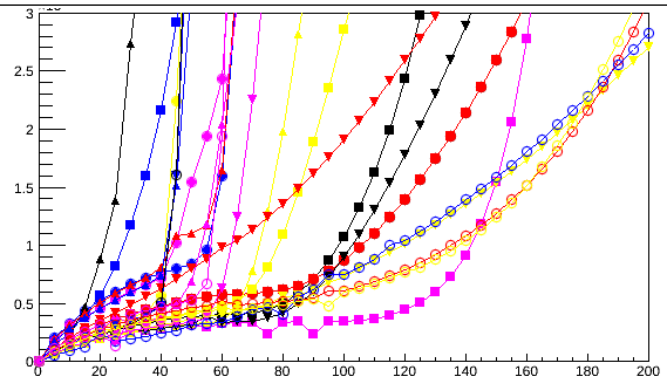


Figure 3.9 IV curves of 25 rejected sensors, showing a fast increase of the leakage current at bias voltages below 200 V. The leakage current [μA] is plotted as a function of the bias voltage [V].

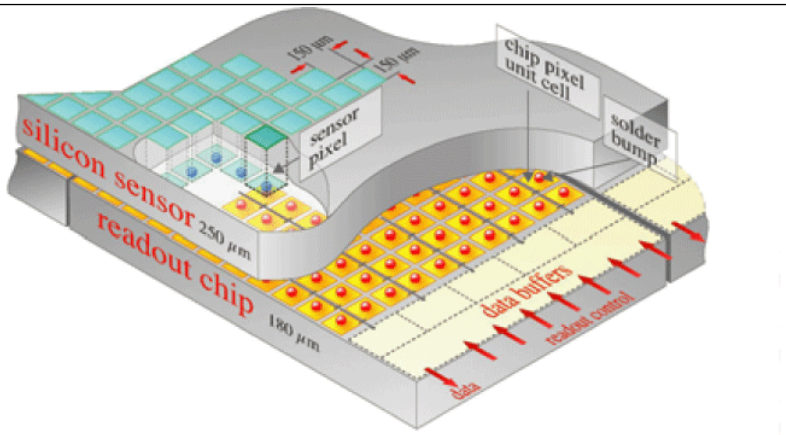


3.3 Read Out Chip

The Read Out Chip (ROC) is an integrated circuit needed to collect, digitize, store and dispatch the pixel signals whose charge exceeds a given threshold. Information attached to each hit are its address, pulse height and time stamp. Each ROC includes 52×80 pixel unit cells, each bump-bonded to a single pixel, and a common periphery hosting control interfaces and data buffers, as sketched in Figure 3.10. Hits are stored temporarily for groups of 160 pixels placed in two neighbouring columns (double column).

ROCs are produced on 8 inches wafers, each of them containing 244 chips, as can be seen in Figure 3.11.

Figure 3.10 Scheme of the pixel silicon sensor bump bonded to the read out chip. Each pixel on the sensor is connected to the corresponding pixel unit cell on the read out chip. The ROC periphery hosts data buffers. Dimensions indicated on this picture are not the ones chosen in the final design.



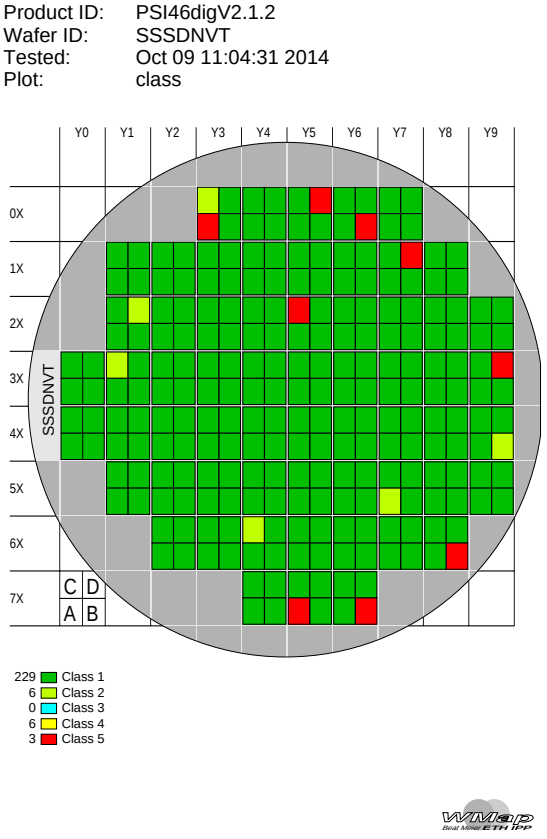
In order to reduce data losses that would occur when operating the present ROC at high LHC luminosity, an upgraded ROC has been developed. It maintains the same architecture, with enhanced performances thanks to the following improvements:

- faster readout thanks to a digital pixel charge readout
- increased number of storage buffer cells
- addition of one buffer stage on the ROC level
- lower charge threshold with which the ROC can be operated, thanks to the reduced noise provided by the digital readout

The ROC has 12 DAC parameters and 2 registers for each pixel, plus four trim bits to equalize pixel thresholds. Single noisy pixels can be masked and consequently excluded from the readout. Moreover, a test circuit included on every pixel allows to qualify the readout chain by directly applying an adjustable calibration signal, hence without an external source. The amplitude of the injected signal is controlled by the `Vcal` DAC parameter, its timing by the `CalDel`. A high or low range of the signal amplitude can be selected with the help of a control register.

The signal from the sensor (or from the internal calibration mechanism) is preamplified and shaped. Then it is compared to a reference voltage, set on ROC basis by the `VthrComp` DAC parameter. If the signal amplitude exceeds the reference voltage, the signal pulse height is read and if the pixel is enabled, the double column periphery is informed. The signal is then stored in the buffer cells and made available for readout by the external circuitery.

Figure 3.11 Location of the 244 chips on the 8 inches ROC wafer, before dicing. Colours indicate the ROC quality: green class 1 are good ROC that can be used for production, while the others have either defective PUC or signal and power connection and will be rejected.



3.4 Bare Modules

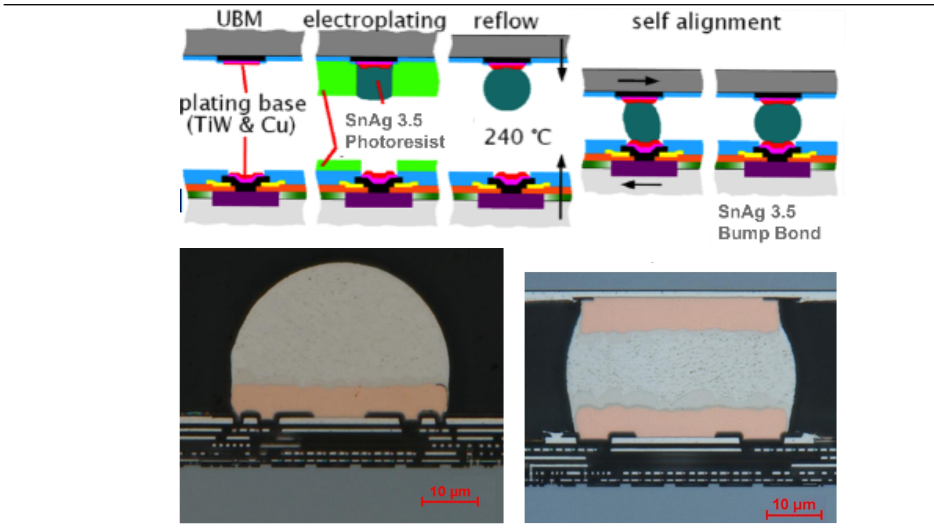
The sensor connected to the ROCs and after gluing the base strips is called bare module, as it is still missing the HDI and readout cables. Each bare module is qualified in order to prevent failures in the interconnection (Bump Bonding) process, and before further assembly steps are involved. In the following paragraphs details of the Bare Module construction and qualification procedure are discussed.

3.4.1 Bump Bonding Process

The first step of assembling a module consists in connecting each pixel to its pixel unit cell located on the ROC. Such connection is established by depositing and soldering bumps between two metal pads, one located on the ROC and one on the corresponding sensor PUC, giving that the connection are uniformly distributed with high density (approximately 5000 in a $\sim$1 cm^2 area). For the Layer 3 modules built by the Italian

consortium, soldering balls are made of an alloy of tin and silver, with a nominal diameter of 36 μm. The bump bonding process is organized in several steps, summarized in Figure 3.12. First, a plating base of titanium and copper is deposited on the under-bump metal pads, both on the sensor and ROC sides. This allows joint solder and protects the silicon parts from metal migration. With the help of a lithographic mask, a layer of photoresist is deposited on the sensor and on the ROC, excluding the bump bonding site. On the ROC side, this empty space is filled with a cylinder of silver-tin, which will provide the bump bonding solder bump. After removing the photoresist layer, the ROC is heated up to 240° C. Thanks to the effect of the superficial tension, the soldering material gains an almost spherical shape. Then, ROCs are thermo-compressed on the sensor and a first soldering is established. After all 16 ROCs have undergone this procedure, the module is heated up to 240° C for a second time, using a particular thermal profile to finalize the solder. This step also allows soldering balls to partially recover their spherical shape, that could have been spoiled by the compression. As bump bonding balls are now effectively connected to the sensor on one side and to the ROC on the other, their reshape compensates for possible small misalignments between the ROC and the sensor, resulting in an advantageous self alignment. The bottom pictures of Figure 3.12 shows a transverse section of the tin-silver bump solder balls, placed on the ROC surface before being connected to the sensor (left) and after being soldered to the sensor (right), seen with a SEM microscope. The robustness of the soldering obtained with the above described process can be estimated from pull tests: a vertical force of about 140 N must be applied to destroy the connection and separate the ROC from the sensor.

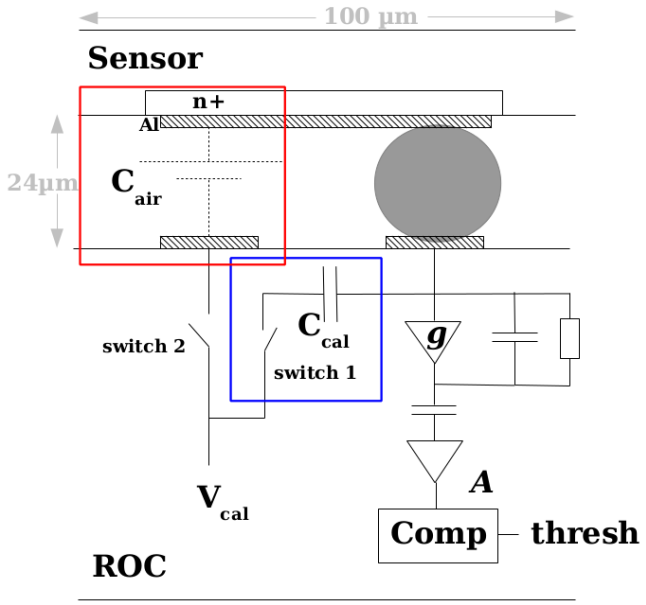
Figure 3.12 Top: Schematic view of the steps of the tin-silver bump bonding process. Bottom: Transverse section of tin-silver solder balls deposited on the ROC surface, before (left) and after (right) being connected to the sensor, as seen with a SEM microscope.



3.4.2 Tests of Bare Modules

Unlike the silicon sensor, the sixteen ROCs are single pieces so that one (or more) of them can be replaced if needed. This procedure is called reworking. For this reason, the bare module is tested to verify the quality of the bump bonding connections and the operation of the chip. As the module is still bare, that is without HDI and cables, a specific probe card is needed to connect the module to a digital test board. It is then connected to the computer that runs the test software. Bare module tests are performed on a single ROC basis, the probe card having 35 needles spaced out by 150 μm to connect the ROC readout pads. Once one ROC has been tested, the probe card is moved to the next one, and the procedure is repeated 16 times per module. Figure 3.13 shows a schematic view of the pixel readout circuitery relevant for the tests described in the following paragraphs.

Figure 3.13 Schematic view of the pixel readout circuit relevant for bump bonding tests. The bump dimension depends on the technology chosen for bump bonding. Modules built by the italian INFN consortium have 36μm SnAg bumps.



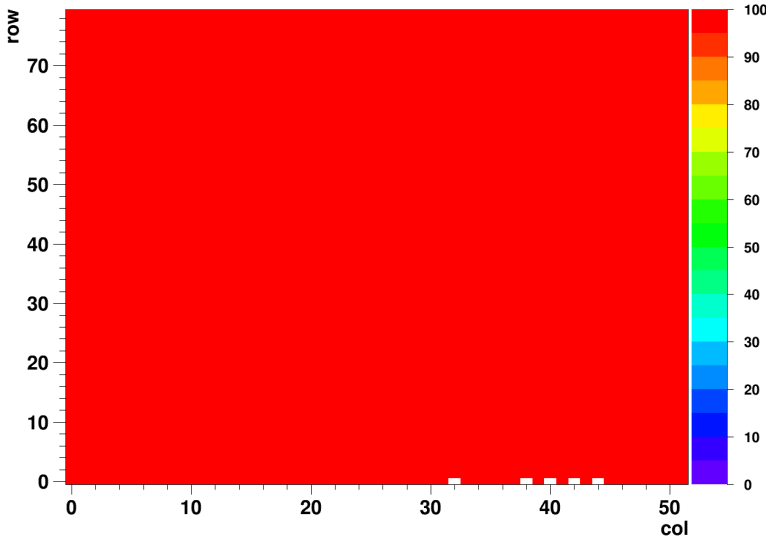
Pixel Alive Test

After finding a convenient working point and adjusting DAC parameters accordingly, it is checked that every pixel can be addressed, read out and masked. A mask bit associated to each pixel allows to disable the comparator in Figure 3.13, therefore suppressing all hits in this pixel. This functionality is very important, since a noisy pixel can fill all the buffers in the double column periphery, preventing it to work properly. The mask test enables the mask bit of a given pixel and tries to read it out. If it is still possible and a pixel hit is found, the mask bit is defective.

The next test, called Pixel Alive test, consists in sending a calibration signal to each pixel unit cell of the ROC and try to read it out. While doing this, the switch 1 in Figure 3.13 is closed and the second switch is opened, so that the read out of the signal is at first order independent from the connection between the sensor and the chip, allowing to probe the correct operation of the ROC itself.

Figure 3.14 shows the result of the Pixel Alive test for one ROC, reporting the number of responses of each pixel unit cell, out of 100 input signals. A high pulse height for the injected signal is chosen, such that a missing read out is not due to an below threshold signal. Not responding pixels are classified as dead pixels.

Figure 3.14 Example of a Pixel Alive test output for one ROC. The map shows the number of responses of each pixel unit cell, out of 100 triggers. Five dead pixels out of 4160 are found in the first row.



Bump Bonding Test

This test aims at checking the quality of the bump bonding connection. After closing switch 2 in Figure 3.13, a calibrate signal V_{cal} is injected to a pad on the ROC surface, inducing a charge in the sensor that mimics a hit. In principle, the hit is detected if and only if the bump is present. However, this is only an ideal situation: the bump can in fact be just poorly connected (but not completely missing), and for large enough injected signals a hit can be triggered although the bump is missing. Such behaviour is attributed to cross talk between the calibration voltage line and the preamplifier, due to a parasitic capacitance. To account for the above mentioned effects, the following algorithm for the bump bonding test is used. First, the threshold for the calibration signal amplitude, when it is injected through the sensor is determined. Then, the threshold for the calibration signal to give a parasitic cross-talk hit is measured. It is done opening both switches in Figure 3.13. If the bump is missing, both thresholds are expected to

be very similar, differently from what expected for good bump connections. Therefore, the difference of the two threshold values is used to distinguish between well and bad connected bumps. More in detail, the following procedure is applied. The distribution of thresholds differences is plotted for all pixels in a ROC and fitted with a gaussian shape, providing the mean and root mean square σ of the distribution for that given ROC. Pixels lying more than 5 σ far from the mean of the distribution are recognized to have missing bumps. The top plot in Figure 3.15 shows the distribution of the thresholds difference for one ROC where a missing bump has been found. As the threshold difference is measured in DAC units, missing bumps are the ones with a threshold difference above (and not below) the 5 σ cut value. The bottom plot in Figure 3.15 represents the threshold differences measured for each pixel of the ROC, allowing to identify the location of the defective pixel. This strategy has been validated on specially prepared bare modules, where a few bumps were manually removed.

Non-destructive Bump Height Measurement

Beside testing weather each pixel is electrically connected to the corresponding read out unit, it is also desirable to check the uniformity of the bump solder balls. A way to do it is just cutting the bare module and observe a transverse section with a SEM microscope, as shown in Figure 3.12. This strategy allows to directly observe the position, shape and dimensions of the bump bonds, but implies destroying the module, thus is done only for a few modules.

An alternative non-destructive measurement to evaluate the uniformity of the bumps height can be performed through electrical tests, exploiting the design of the ROC.

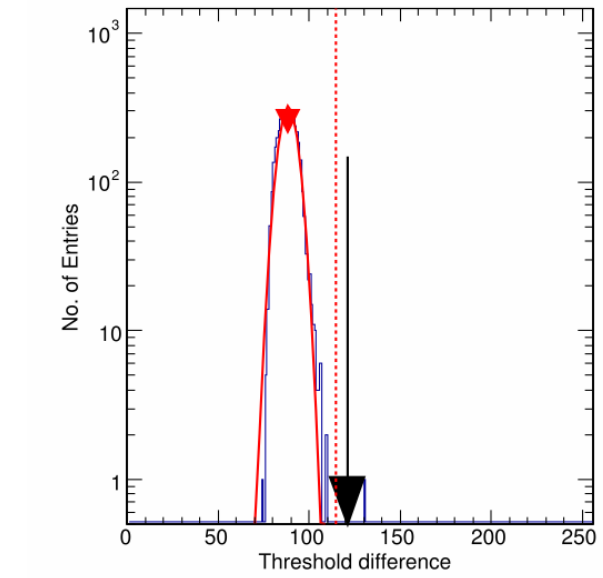
Observing Figure 3.13, it can be seen that a test pulse V_{cal} can be applied in two ways:

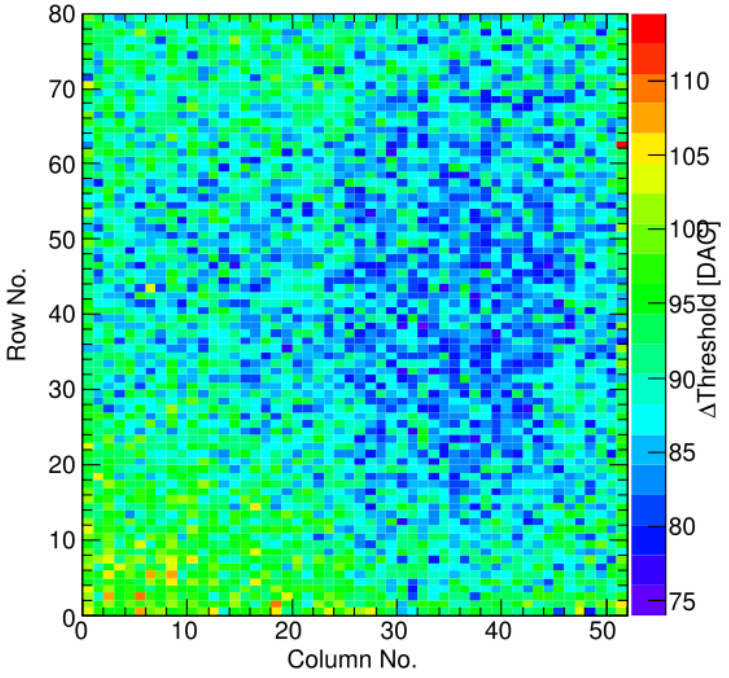
1. directly through the C_{cal} capacitance, closing switch 1, with switch 2 opened. In this configuration, the test pulse is injected and read out through the ROC, without relying on the connection between the chip and the sensor. Therefore, a calibration signal applied this way is labeled with the subscript ROC
2. through the air capacitance C_{air} formed by a metal pad on the ROC and the sensor pad. If a pulse V_{cal} is applied with swicth 2 closed and switch 1 opened, a mirror charge appears on the sensor pad. If the bump bond connection is in place, this mirror charge flows through the bump to the ROC and is read out. The calibration signal is injected following this path is labeled with the subscript SEN, to remind that the output signal is read out from the sensor through the bump.

The output signal is sent to a preamplifier with gain factor g, hence its amplitude A is proportional to the induced charge Q, the latter depending on the capacitance C through which the calibration pulse V_{cal} is applied:

$$A = g \cdot C \cdot V_{cal} \tag{3.2}$$

Figure 3.15 Top: distribution of the threshold difference in DAC units, for pixels in one ROC. The bulk of the distribution is fitted with a gaussian shape with parameters μ and σ. The cut value for missing bumps is chosen to be $\mu + 5\sigma$ and is represented by the red dotted line. One missing bump is found in this ROC. Bottom: map of the thresold difference for pixels in the ROC, allowing to identify the position of the defective pixel, that is in row 63, column 52.





Hence, when the signal V_{cal} is injected through the internal capacitance C_{cal} (case 1)

$$A_1 = g \cdot C_{cal} \cdot V_{cal,ROC} \tag{3.3}$$

while if it is applied through the air capacitance C_{air} (case 2)

$$A_2 = g \cdot C_{air} \cdot V_{cal,SEN} \tag{3.4}$$

The signal is then sent to a comparator, and is read out only if its pulse height is higher than the comparator threshold V_{thr}. Such threshold can differ from pixel to pixel, but does not depend on how the calibration signal is applied (and read out). Instead, the amplitude of the calibration signal needed to have a pulse height greater than the comparator threshold, which is indicated with V_{cal}^{thr} is different in the two cases:

$$\begin{aligned} V_{thr} &= g \cdot C_{cal} \cdot V_{cal,ROC}^{thr} \\ V_{thr} &= g \cdot C_{air} \cdot V_{cal,SEN}^{thr} \end{aligned} \tag{3.5}$$

Equating the two expressions in Equation 3.5 leads to the relation between the capacitances and the threshold values for the calibration signal:

$$V_{cal,SEN}^{thr} / V_{cal,ROC}^{thr} = C_{cal} / C_{air} \tag{3.6}$$

While the C_{cal} capacitance is a fixed construction parameter of the chip itself, the air capacitance depends on the gap between the metal pads (one on the ROC surface and one on the sensor) that is determined by the height of the bump solder ball. In the ideal case of a plane parallel sides capacitor, the capacitance is inversely proportional to the distance between the pads. Taking it into account in Equation 3.6 allows to obtain a relation between the threshold ratio and the bump height h_{bump}

$$V_{cal,SEN}^{thr} / V_{cal,ROC}^{thr} \propto 1/C_{air} \propto h_{bump} \tag{3.7}$$

Equation 3.7 means that the thresholds ratio is proportional to the bump height. Therefore, measuring the threshold voltages for the calibration signal applied in the two configurations and taking their ratio for each pixel gives an estimate of the bump height uniformity.

The threshold voltage $V_{cal,thr}$ for each pixel can be measured counting the number of responses as a function of the calibration pulse height V_{cal}. If the system were noise free, the distribution of the number of hits as a function of the applied signal would be a single step function, the discontinuity point identifying the voltage threshold. In a real system, noise causes charge fluctuations. For Gaussian fluctuations, the efficiency curve as a function of the pulse height can be described by an error function, whose width

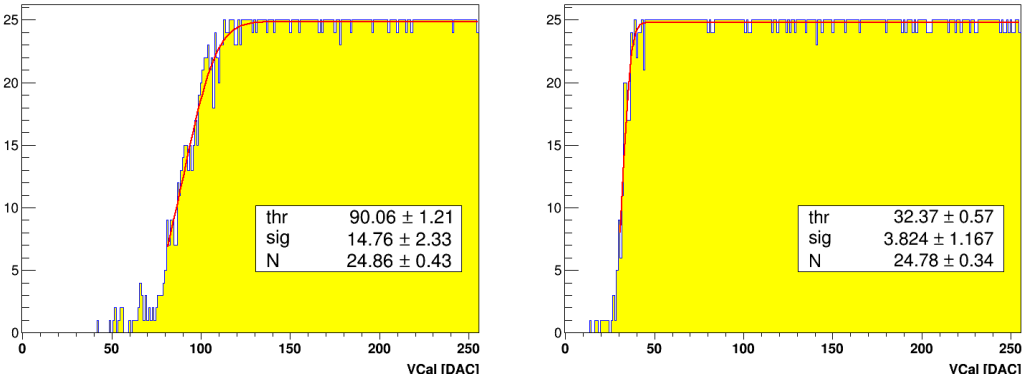
gives an estimate of the noise. The threshold is defined as the voltage corresponding to the 50% efficiency point.

Figure 3.16 shows the number of responses out of 25 triggers as a function of the calibration signal V_{cal}, applied in the two ways explained before, for a randomly chosen pixel. A χ^2 fit with the function

$$f(x) = \frac{N}{2}\left[1 + Erf\left(\frac{x - x_{thr}}{\sqrt{2}\sigma}\right)\right] \tag{3.8}$$

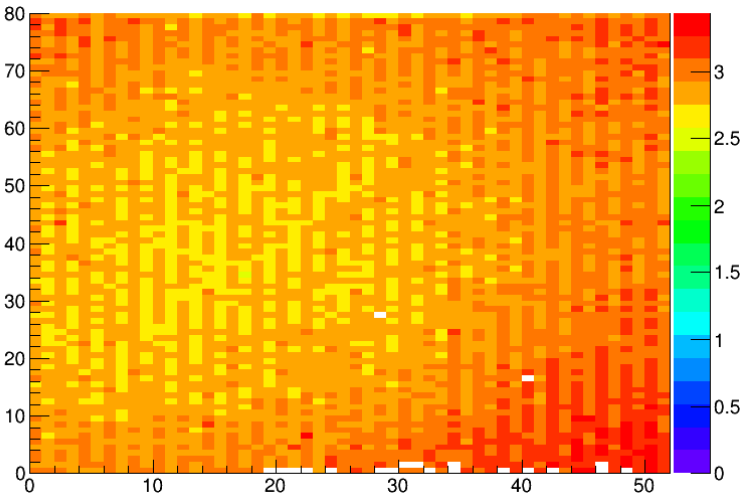
is performed. It returns the best values for the parameters N, x_{thr} and σ, representing the number of triggers, the 50% efficiency point and the curve width respectively. In the plots in Figure 3.16, the best fit is represented by the continuous curve. The threshold values whose ratio is proportional to the bump height are given from the fit parameter x_{thr}.

Figure 3.16 Number of responses out of 25 triggers as a function of the calibration signal amplitude V_{cal}, when it is read out through the sensor (left) or the internal ROC capacitance (right). The red lines represent the best fit curves from which the threshold values are obtained. The V_{cal} pulse height is measured in DAC units from 0 to 255.



This same procedure is repeated for all pixels in a ROC to build a map of the measured threshold ratio values. As an example, Figure 3.17 shows a ratio map for a ROC of one of the tested modules. For some pixels, the ratio value could not be measured as one (or both) of the two thresholds could not be estimated due to poor convergence of the fit or to missing (or very low) response of the pixel, that can be caused by defects in the bump connection or in the read out circuit. The uncertainty on the threshold ratio has been estimated from the average and RMS of the distributions for the thresholds $V^{thr}_{cal,SEN}$ and $V^{thr}_{cal,ROC}$ over one ROC. Typical values are $V^{thr}_{cal,SEN} = 90 \pm 5$ and $V^{thr}_{cal,ROC} = 30 \pm 1$ DAC units, with a threshold ratio ~ 3. Propagating the error, an uncertainty on the threshold ratio of ~ 0.2 is obtained. As can be seen from the map in Figure 3.17, the threshold ratio has a variation of ~ 0.5 over most of the ROC, with an increasing trend when moving out from the center, indicating that the bump height is greater close to the corners. Such behaviour can be due to bowing of the ROC surface caused by thermal stress during the bump bonding procedure. Dedicated measurements have shown that a maximum bow of 1 μm is observed after a thermal cycle up to 250° C. However, variations are small and the uniformity of solder bumps heights is satisfactory.

Figure 3.17 Map of the threshold ratio $V^{thr}_{cal,SEN}/V^{thr}_{cal,ROC}$ for the pixels in a ROC of one of the tested modules.



3.4.3 Bare Modules Grading

According to the results of the tests, a bare module is graded A,B or C in order of decreasing quality. Grading criteria are defined on a single ROC basis, the grade of the whole module being the worst among its 16 ROCs. Several factors are combined to define the quality class of a bare module. First, the IV curve is graded:

A if the current i at a bias voltage of 150 V is smaller than 2 μA and the ratio i(V=150 V)/i(V=100 V) < 2

B if the current i at a bias voltage of 150 V is smaller than 10 μA and the ratio i(V=150 V)/i(V=100 V) < 2

Then, the ROC failures are defined as the sum of the number of dead pixels plus the number of missing bumps found with the Pixel Alive and Bump bonding tests respectively. Moreover, the current flowing in the digital periphery of the ROC is required to be smaller than 64 mA. A bare module is graded

A if its IV curve is graded A and the amount of ROC failures is less than 1%

B if its IV curve is graded B and the amount of ROC failures is less than 4%

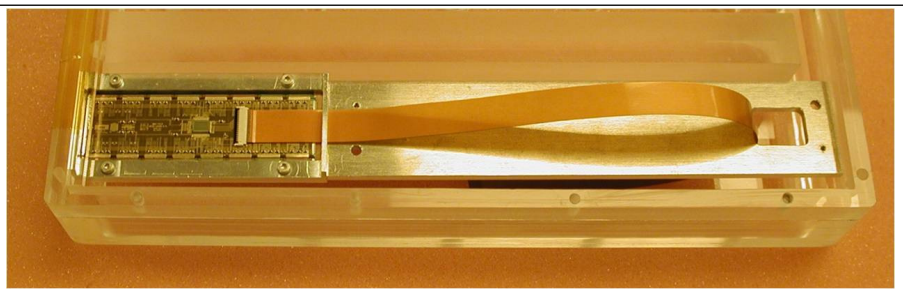
C if the IV curve or the digital current do not fulfill the requirements or if the number of ROC failures is greater than 4%.

A minus sign can be added to the A or B grade meaning that one or more ROCs have been reworked.

3.5 Full Modules

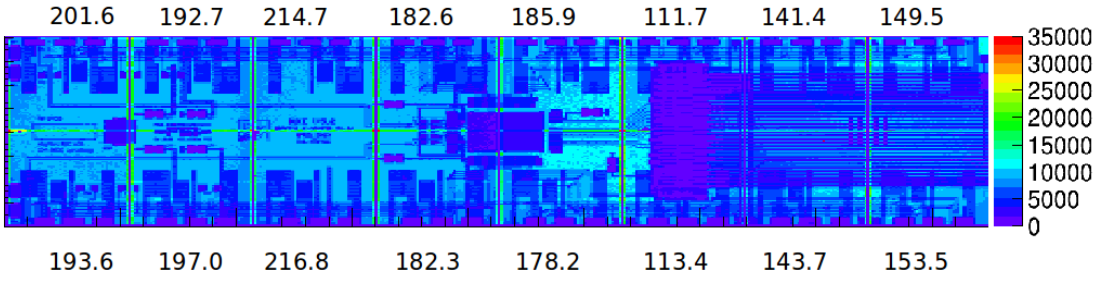
After being tested and qualified, a bare module is glued and wire bonded to the HDI circuit, the TBM chip and the readout cable to obtain a full module. Figure 3.18 shows an assembled full module on its handler.

Figure 3.18 Picture of an assembled full module placed on its handler.



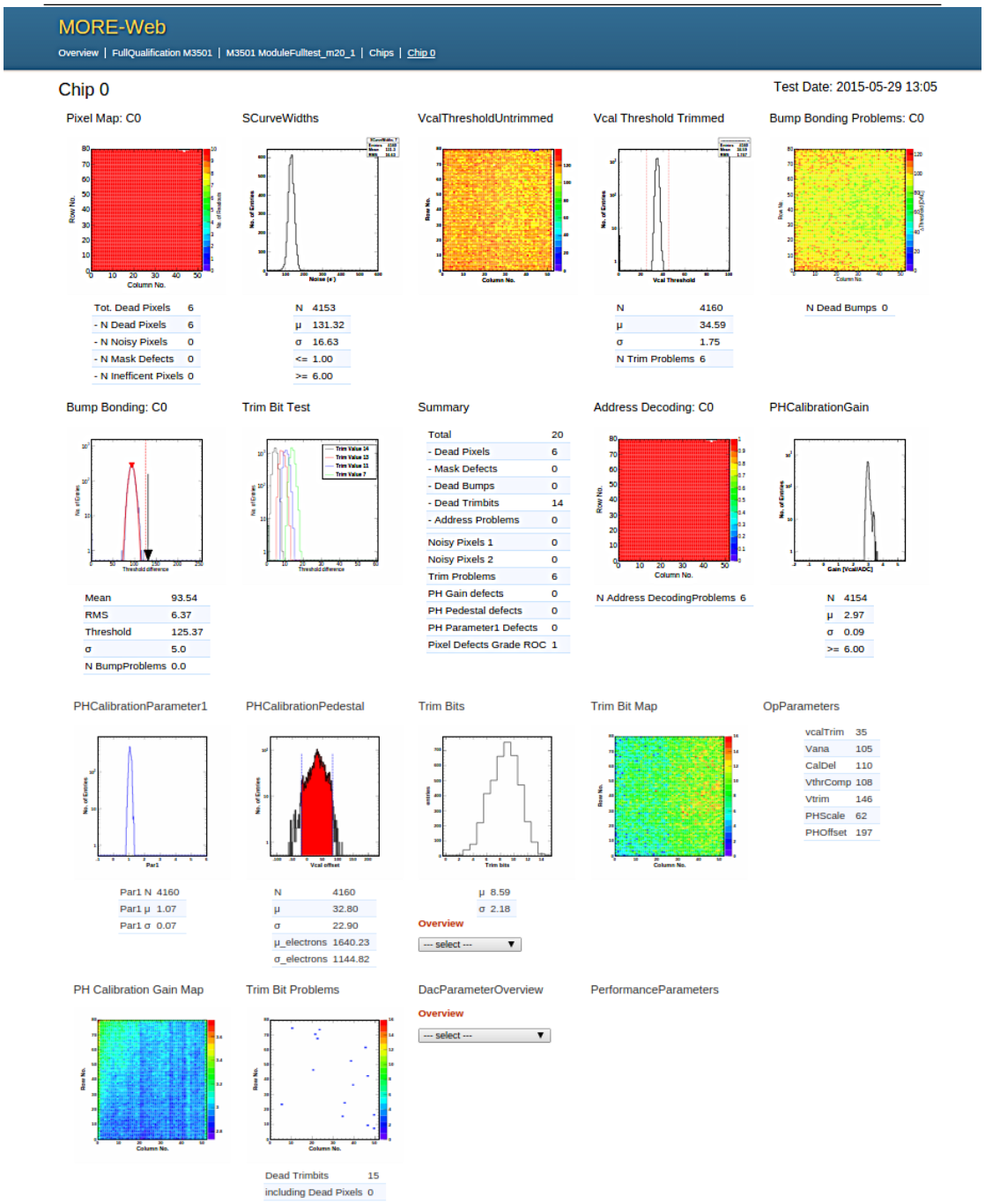
Further functionality tests are then performed, which are similar to the ones already described for the bare module. In addition to that, the full module undergoes thermal cycles in a cold box controlled environment to verify the stability under real operating conditions. The module is cooled from 17° C down to -25° C within less than 15 minutes for ten times and tested again afterwards. After that, X-ray tests are performed to verify that pixels can actually collect charge deposited by an external ionizing source. The module is irradiated with a beam of X-rays with the energy spectrum and intensity of the emitter tube in order to obtain a high statistics hit map for every PUC, as can be seen in Figure 3.19. At this stage, missing or defective bumps can be directly and easily identified, allowing to cross check bump bonding test results performed on the bare module. An energy calibration is then performed using fluorescence X-rays emitted from four different targets that are irradiated by the direct X-ray beam. Finally, four trim bits are used to further equalize the pixel thresholds within each ROC.

Figure 3.19 Hit map of a full module irradiated with a direct beam of X-rays.



As an example, Figure 3.20 shows the overview of results of a full module test for one ROC, displayed through a web interface produced by a dedicated software called MORE-Web (CMS Pixel Detector Module Qualification Result Web Interface).

Figure 3.20 Example of full module test results for one ROC. The overview module test results is displayed as a web page produced by the MORE-Web software.



3.6 Production Data Base

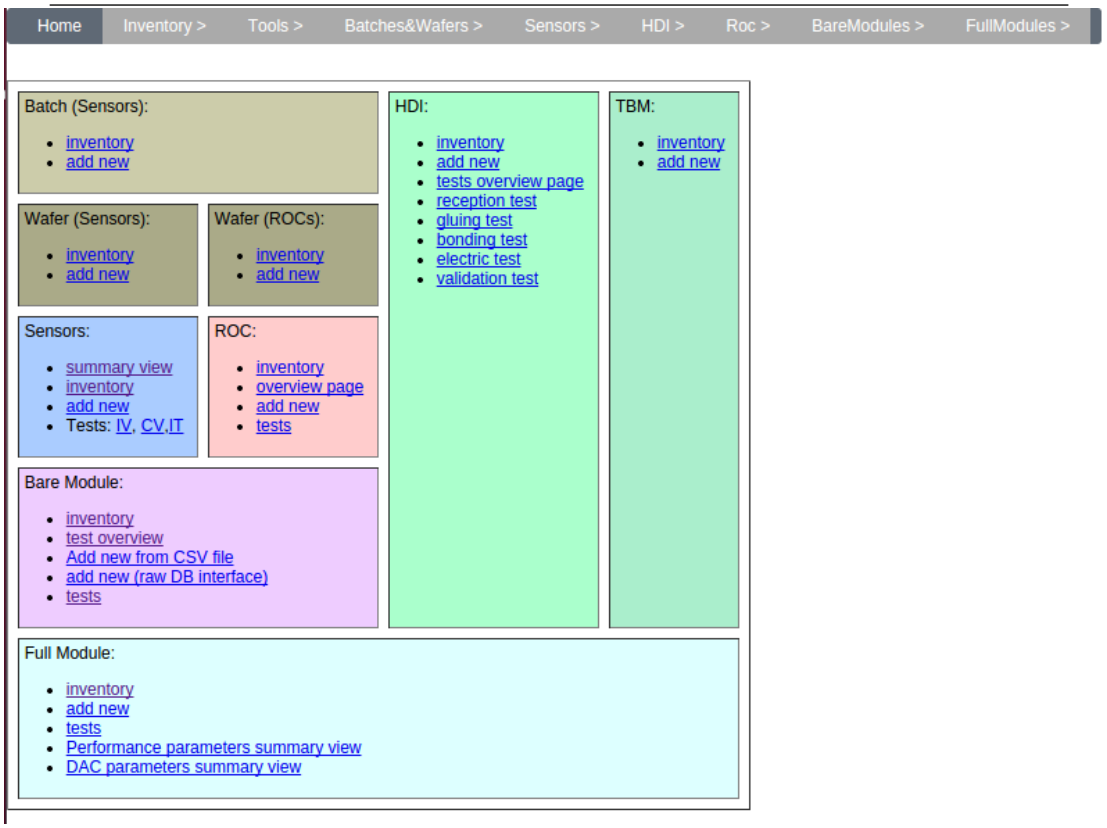
All information about shipping and delivering of module components for the construction of the BPIX tracker and qualification test results are collected and stored in a central production data base.

Each production center is required to upload test results in a standard format, decided according to the kind of test. Usually they are plain text and CSV files summarizing test parameters together with most relevant measurements and defects found. The grading of parts according to established requirements are also given. In addition to that, access to full measurement files is provided.

The data base is organized in several sections, one for each module component. Different objects used to build a module are linked to it, allowing an easy and user-friendly access to the quality assurance tests of single components.

The data base interface consists of a web page that the user can navigate, as can be seen in Figure 3.21. It is updated automatically every 15 minutes, providing a useful tool to monitor the production status and to compare and exchange information among all several centers involved in the BPIX construction.

Figure 3.21 Home page of the BPIX production data base, showing the different sections corresponding to the module components.



Chapter 4

Particle Identification and Reconstruction in CMS

In this Chapter, the main algorithms for object reconstruction and particle identification adopted by the CMS Collaboration are introduced.

4.1 Tracks

In the quasi-uniform magnetic field of the tracker, charged particles follow helical paths and therefore five parameters are needed to define a trajectory. Extraction of these five parameters requires at least three 3D hits, or two 3D hits and a constraint on the origin of the trajectory based on the assumption that the particle originated near the beam spot. The luminous region is related to the dimensions of proton bunches, that have a Gaussian profile in the transverse plane, with σ_x=16 μm and σ_y =100 μm, and are approximately 15 cm long.

Reconstructing the trajectories of charged particles starting from the collection of hits in the tracker is a computationally challenging task. The tracking software of CMS is based on an iterative structure. Iteration 0, the source of most reconstructed tracks, is designed for prompt tracks (originating near the pp interaction point) with $p_T > 0.8$ GeV and having three pixel hits. Iteration 1 is used to recover prompt tracks that have only two pixel hits. Iteration 2 is configured to find low-p$_T$ prompt tracks. Iterations 3-5 are intended to find tracks that originate outside the beam spot and to recover tracks not found in the previous iterations. At the beginning of each iteration, hits associated with high-purity tracks found in previous iterations are excluded from consideration. Each iteration proceeds in the four following steps, briefly described below: seed generation, track finding, track fitting and track selection [47].

Seed Generation Seeds define the starting trajectory parameters and associated uncertainties of potential tracks. They are constructed from hits in the pixel tracker

and track candidates are built outwards. This approach benefits from the low channel occupancy and 3D spatial measurement in pixel sensors, providing more constraints and better estimates of trajectory parameters. Moreover, generating seeds in the inner tracker leads to a higher reconstruction efficiency, allowing to detect particles that do not reach the outer tracker, e.g. electrons losing a significant part of their energy traversing the tracker, pions undergoing nuclear interactions or low-momentum tracks trapped by the strong magnetic field. Seeds are mainly triplets of pixel hits. Nevertheless, other trajectory seeds are needed to compensate for inefficiencies in the pixel detector (from gaps in coverage, non-functioning modules, and saturation of the readout), and partially to reconstruct particles not produced directly at the pp collision point.

Track Finding It is based on the Kalman filter method. The filter begins with a coarse estimate of the track parameters provided by the trajectory seed, and then builds track candidates by adding compatible hits from successive detector layers, updating the parameters at each layer if more than one compatible hit is found. The information needed at each layer includes the location and uncertainty of the detected hits, as well as the amount of material crossed, which is used to estimate the effects of multiple Coulomb scattering and energy loss. The standard termination conditions are if a track reaches the end of the tracker or contains too many missing hits, or if its momentum drops below a user specified value. When the search for hits in the outward direction reveals a minimum number of valid hits, an inward search is initiated for additional hits. The track of a single charged particle can be reconstructed more than once, either starting from different seeds, or when a given seed develops into more than one track candidate. To remedy this feature, a trajectory cleaner is applied after all the track candidates in a given iteration have been found. Each pair of tracks undergo a trajectory cleaner process, that identifies tracks sharing more than a given fraction of hits and discards the track with the fewest hits (or with the largest χ^2, when tracks have the same number of hits).

Track Fitting For each trajectory, the track-finding stage yields a collection of hits and an estimate of the track parameters. However, the full information about the trajectory is only available at the final hit of the trajectory (when all hits are known). Furthermore, the estimate can be biased by constraints, such as a beam spot constraint applied to the trajectory during the seeding stage. The trajectory is therefore refitted using a Kalman filter and smoother. To obtain the best precision, this filtering and smoothing procedure takes into account not only the effect of material, but also inhomogeneities of the magnetic field especially in the $|\eta| > 1$ region. Estimates of the track trajectory at any other points, such as the point of closest approach to the beam-line, can be obtained by extrapolating the trajectory evaluated at the nearest hit to that very point, using the same fitting technique. After filtering and smoothing, a search is made for spurious hits incorrectly associated to the track. Such hits can be related to an otherwise well-defined track, to nearby tracks, to delta rays or to electronic noise.

Track Selection In a typical LHC event, the track-finding procedure described above yields a significant fraction of tracks not corresponding to a charged particle, the so called fake tracks. The fraction of fake tracks can be substantially reduced through quality requirements. Tracks are selected on the basis of the number of layers that have hits, of the goodness of the χ^2 and of the compatibility with a primary interaction vertex. If several primary vertices are present in the event, as often happens due to pileup, all are considered.

Track Reconstruction Efficiency and Fake Rate

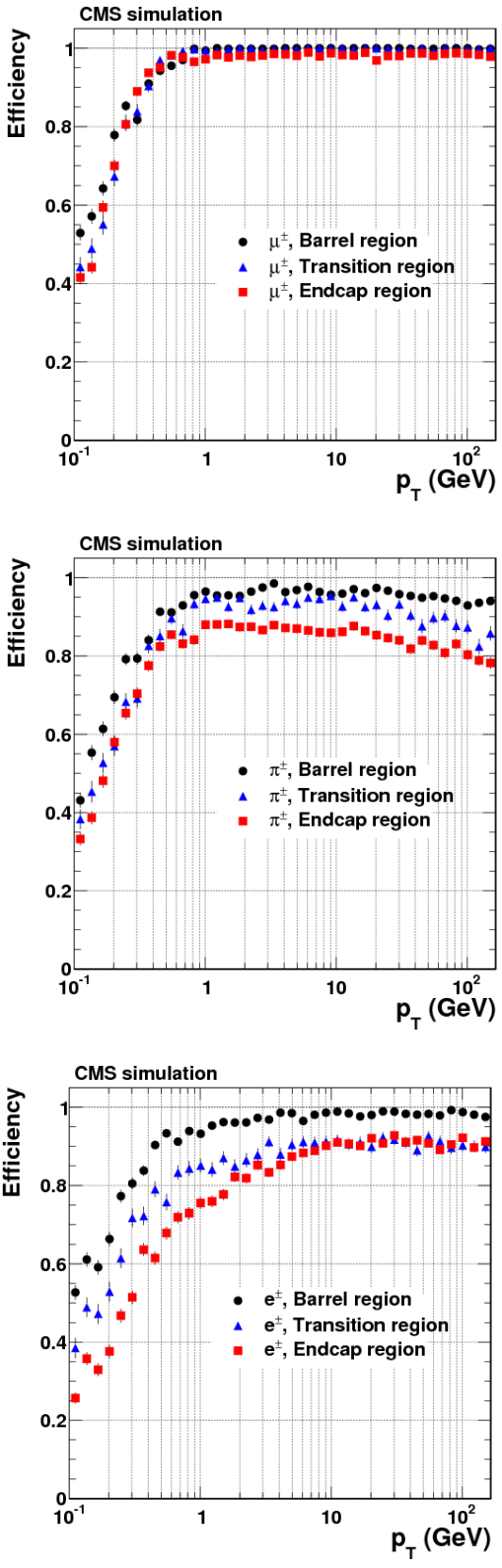
For simulated samples, the tracking efficiency is defined as the fraction of simulated charged particles that can be associated with corresponding reconstructed tracks, where the association criterion is the one described at the beginning of this Section. Such definition of efficiency depends not only on the quality of the track-finding algorithm, but also upon the intrinsic properties of the tracker, such as its geometrical acceptance and material content. Using the same association criterion used for the efficiency, the fake rate is defined as the fraction of reconstructed tracks that are not matched to any simulated particle.

Plots in Figure 4.1 shows the track finding efficiency for single isolated muons, pions and electrons as a function of the track p_T in barrel ($0 < |\eta| < 0.9$), endcap ($1.4 < |\eta| < 2.5$) and in the intermediate transition ($0.9 < |\eta| < 1.4$) region.

Muons are reconstructed better than any other charged particle in the tracker, as they mainly interact with the silicon detector through ionization of the medium and, unlike electrons, their energy loss through bremsstrahlung is negligible. Muons therefore tend to cross the entire volume of the tracking system, producing detectable hits in several sensitive layers of the apparatus. Finally, muon trajectories are altered almost exclusively by Coulomb scattering and energy loss, whose effects are well taken into account by the Kalman filter. For isolated muons with $1 < p_T < 100$ GeV, the fake rate is completely negligible, the tracking efficiency is larger than 99% over the full tracker acceptance, and does not depend on p_T.

Charged pions are also subject to elastic and inelastic nuclear interactions. Elastic nuclear interactions introduce long tails in the distribution of the scattering angle that are not taken into account during the track finding. As a result, the formation of a track can be interrupted if a hadron undergoes a large-angle elastic nuclear scattering. Inelastic nuclear interactions are the main source of tracking inefficiency for hadrons, particularly in those regions of the tracker with large material content. Depending on η, up to 20 % of the simulated pions are not reconstructed. This effect is most significant for hadrons with $p_T \leq 700$ MeV, because of the larger cross sections for nuclear interactions at low energies. The tracking efficiency is also affected, along with the fake rate by the secondary particles produced in inelastic processes, particularly for high momentum pions. While the fake rate is generally smaller than $2-3\%$ for tracks reconstructed in a

Figure 4.1 Track reconstruction efficiencies for single isolated muons (top), pions (middle) and electrons (bottom) as a function of p_T. The barrel, transition and endcap regions are defined by the η intervals $[0, 0.9]$, $[0.9, 1.4]$, and $[1.4, 2.5]$ respectively [47].



sample of single pions with $p_T = 1$ or 10 GeV, in a sample of single pions with a p_T of 100 GeV, the fake rate increases up to 15 % in the transition region between the barrel and the endcaps.

Electrons lose a large fraction of their energy via bremsstrahlung radiation before they reach the outer layers of the silicon tracker. Hence, the number of hits assigned to the track can be significantly reduced. Moreover, if a radiated photon converts to an electron-positron pair or induces an electromagnetic shower, the track finder can assign a mixture of hits from the primary electron and from the secondary particles to a single track. This reduces tracking efficiency, increases fake rate and is the main source of charge misidentification for electrons. In the barrel, the efficiency for electrons exceeds 90% for $p_T > 0.4$ GeV, and the fake rate is very small. The performance is significantly worse in the endcap and barrel-endcap transition regions, because of the larger amount of traversed material.

4.2 Muons

Muon reconstruction in CMS is performed with the silicon tracker at the heart of the detector, and with up to four stations of gas-ionization muon detectors installed outside the solenoid and sandwiched between steel layers serving both as hadron absorbers and as a return yoke for the magnetic field. The muon system [58] covers the pseudorapidity region $|\eta| < 2.4$ and performs three main tasks: triggering on muons, identifying muons, and supporting the CMS tracker in measuring the momentum and charge of high momentum muons.

Muon tracks are first reconstructed independently in the silicon tracker and in the muon spectrometer. The resulting tracks are called tracker track and standalone-muon track respectively. Based on these, two reconstruction approaches [59] are used:

- global muon reconstruction (*outside-in*): starting from a standalone muon a matching tracker track is found and a global-muon track is fitted combining hits from the tracker track and standalone-muon track. At large transverse momentum ($p_T > 200$ GeV), the global-muon fit can improve the momentum resolution compared to the tracker-only fit.
- tracker muon reconstruction (*inside-out*): all tracker tracks with $p_T > 0.5$ GeV and $p > 2.5$ GeV are considered as possible muon candidates and are extrapolated to the muon system, taking into account the expected energy loss and the uncertainty due to multiple scattering. If at least one muon segment (i.e. a short track stub made of DT or CSC hits) matches the extrapolated track, the corresponding tracker track qualifies as a tracker-muon track. At low momentum (roughly $p < 5$ GeV) this approach is more efficient than the global muon reconstruction, since it requires only a single muon segment in the muon system, whereas global

muon reconstruction is designed to have high efficiency for muons penetrating through more than one muon station.

The majority of muons from collisions (with sufficient momentum) are reconstructed either as a global muon or a tracker muon, or very often as both. However, if both approaches fail and only a standalone-muon track is found, this leads to a third category of muon candidates:

- Standalone-muon track only: this occurs only for about 1% of muons from collisions, thanks to the high tracker-track efficiency. On the other hand, the acceptance of this type of muon track for cosmic-ray muons is a factor 10^2 to 10^3 larger, thus leading to a collision muon to cosmic-ray muon ratio that is a factor 10^4 to 10^5 smaller than for the previous two muon categories.

The combination of different algorithms provides a robust and efficient muon reconstruction. The desired balance between identification efficiency and purity depending on the physics analysis is achieved by applying a selection based on muon identification variables. Standard selections are provided, such as:

- soft muon selection, optimized for low momentum ($p_T < 10$ GeV), requiring the candidate to be a tracker muon with the additional constraint that a segment is matched in both x and y coordinates with the extrapolated tracker track.
- tight muon selection, mainly used in electroweak analysis, applies more stringent requirements on the muon candidate. It must be reconstructed outside-in as a global muon, moreover the following additional requirements must be satisfied: normalized χ^2 of the global muon track fit < 10; at least one muon chamber hit included in the final track fit; matched to muon segments in at least two muon stations (this implies that the muon is also reconstructed inside-out as a tracker muon); and its corresponding tracker track must have more than 10 silicon tracker hits (including at least one pixel hit) and a transverse impact parameter $|d_{xy}| < 2$ mm with respect to the primary vertex. With this selection, the rate of muons from decays-in-flight is significantly reduced.

Muon Reconstruction Efficiency and Misidentification Rate

The efficiency of muon reconstruction [59] can be estimated not only from simulated events but also from data, applying a "tag-and-probe" technique to muons from the decays of J/Ψ and Z resonances. Events with two muons whose invariant mass is compatible with the J/Ψ (or Z) mass are considered. They are selected with tight selection requirements on one muon (the tag muon), and with a loose selection on the other muon (the probe muon), so that the efficiency measurement is not biased by the probe

muon selection. The fraction of probe muons passing the selection under study gives an estimate of its efficiency.

The plateau of the efficiency for soft muons is reached at $p_T \sim 4(6)$ GeV in the barrel (endcaps), while for tight muons it is reached at $p_T \sim 10$ GeV in both regions. Plateau efficiencies for the two muon categories are above 96% both in the barrel and endcap regions, exceeding 99% for soft muons from Z decays.

The probability of misidentification of prompt muons arises mostly from hadrons, and has three main contributions:

- decay in flight of hadrons
- punch-through, due to hadronic showers that are not fully contained in the hadron calorimeter and reach the muon detector
- random matching between the hadron track in the inner tracker and a track stub in the muon system from one of the other tracks in the jet that may be due to a muon.

Overall, the probability to misidentify a hadron as a muon is the largest for soft muons, and drops significantly for tight muons. For pions and kaons, the misidentification probabilities are below 1% for all muon selections and decrease at $p \geq 10 - 15$ GeV due to fewer of the hadrons decaying to muons within the detector volume. For protons, the probability to be identified as a muon slowly increases with momentum but remains low in the accessible momentum range, which confirms that punch-through is small and that at low momentum the main reason for misidentification of pions and kaons is decays in flight.

4.3 Photons

Photons are reconstructed [60] by clustering energy deposits in the ECAL crystals. Due to the magnetic field, electromagnetic showers are spread in the ϕ direction. Clustering algorithms, developed to group ECAL crystals together, start from a seed, that is a crystal whose signal corresponds to a transverse energy greater than those from neighbour crystals, and above a given threshold. In the barrel, where the crystals are arranged in a (η, ϕ) grid, clusters have a fixed width of 5 crystals in the η direction, while on the ϕ direction strips of 5 crystals are added if their total energy is above a defined threshold. In the endcap region, clustering proceeds by adding 5×5 crystal matrices if their are close enough to the seed crystal, again taking into account the ϕ spread.

In order to improve energy resolution, the initial sum of energy deposits is corrected to account for shower containment in the clustered crystals and for shower losses of photons that convert before reaching the calorimeter.

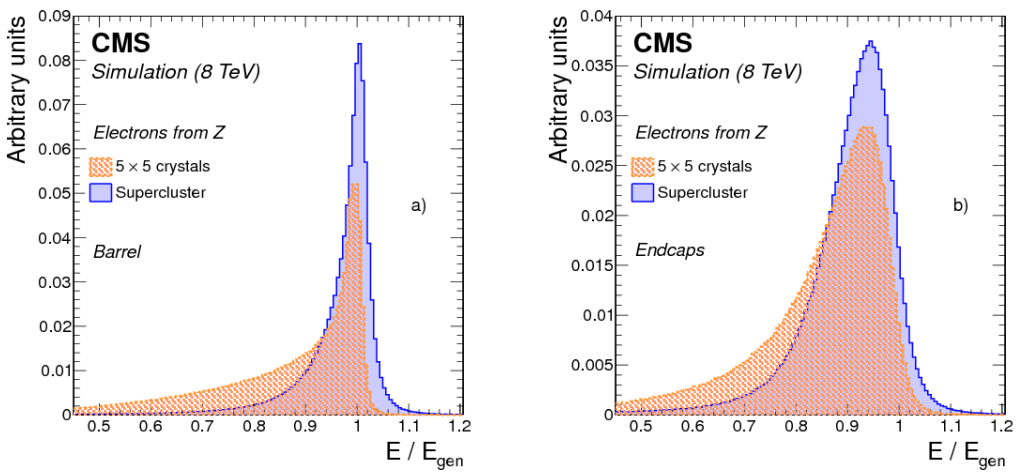
Photon identification is mainly based on shower-shape and isolation variables.

4.4 Electrons

Electrons are reconstructed by associating a track in the silicon tracker with an energy cluster in the ECAL. The energy spread over crystals is small when losses due to bremsstrahlung before reaching the calorimeter are small. A test beam with 120 GeV electrons that impinge on the center of a crystal has shown that about 97% of the energy is deposited in a 5×5 cluster. However, an electron produced within the CMS detector and traveling through the least amount of material ($|\eta| \sim 0$) radiates away an average of 33% of its energy before reaching the ECAL. This number increases up to 86% for the largest depth of material traversed, for $|\eta| \sim 1.4$.

An accurate measurement of the electron initial energy requires the collection of energy of radiated photons that are spread along the ϕ direction. This spread effect is due to the bending of electron trajectories caused by the magnetic field. The spread along the η direction is negligible, with the exception of very low momentum electrons ($p \leq 5$ GeV). Consequently, clustering algorithms search for energy deposits in a small window in η and an extended window in ϕ. They start from a crystal identified as a seed, having an energy deposit above a fixed threshold and greater than its neighbour crystals. In the barrel region, arrays of 5×1 crystals in $\eta \times \phi$ are added around the seed crystal, in both directions of ϕ, if their energies exceed a minimum threshold. Adjacent arrays are grouped into clusters, that are merged to form a supercluster, provided that there is a seed-array with an energy greater than a given threshold. In the endcap region, where crystals are not arranged into a (η, ϕ) grid, the clustering proceeds through merging 5×5 crystal matrices to form superclusters. The supercluster energy is calculated as the sum of its clusters energies, and its position is the mean cluster position, weighted over the energy. Figure 4.2 compares the ratio of reconstructed over generated energy for simulated electrons from Z boson decays, in the barrel and in the endcaps, when using 5×5 crystals and superclusters.

Figure 4.2 Distributions of the ratio of reconstructed over generated energy for simulated electrons from Z boson decays in the barrel (left) and in the endcaps (right), for energies reconstructed using superclustering and a matrix of 5×5 crystals [61].

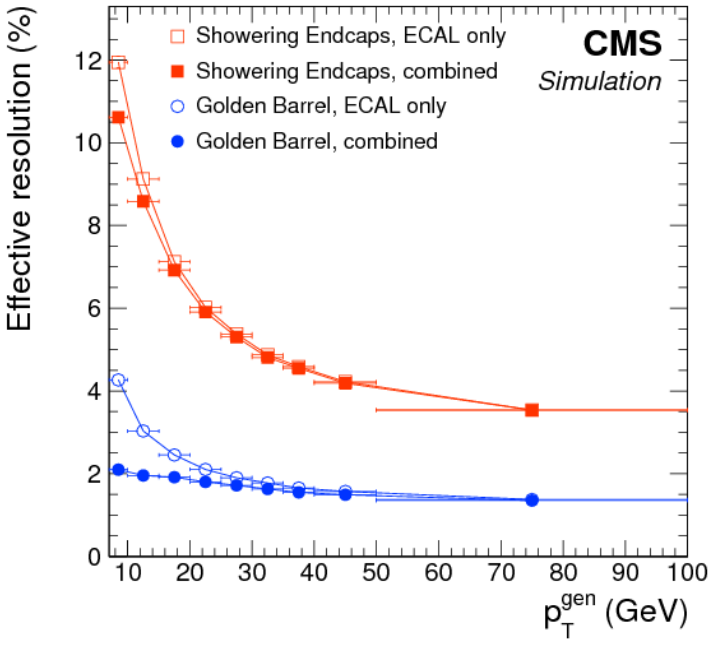


Electrons are classified in the following categories, depending on the fraction of energy loss due to bremsstrahlung in the tracker:

- "golden" electrons, losing only a small fraction of their energy before reaching the ECAL
- "big-brem" electrons, radiating a large amount of their energy in a single step along their trajectory
- "showering" electrons, having large energy losses all along their trajectory

Figure 4.3 shows the transverse momentum resolution estimated with the supercluster energy only and when combining the latter measurement with momentum measurement provided by the tracker. Golden electrons in the barrel and showering electrons in the endcaps are compared.

Figure 4.3 Resolution on the electron transverse momentum as a function of the generated electron p_T, for showering electrons in the endcaps (red) and golden electrons in the barrel region (blue). Open points refer to the p_T estimate using supercluster energy information only, while solid points are relative to the p_T estimate when combining the supercluster energy with the momentum measurement provided by the tracker [61].



4.5 Jets

Almost every process of interest at the LHC contains quarks or gluons in the final state. Partons can not be observed directly, but due to color confinement they fragment into hadrons. Therefore, the experimental evidence of a quark or a gluon is the so called *jet*, that is the collimated shower of particles originating from the hadronization of a parton.

The hadronic calorimeter is the main sub-detector for jet identification and reconstruction, it measures the energy deposited by the parton shower in each crystal. Energy deposits in crystals are then clustered according to given algorithms in order to reconstruct calorimeter jets (*CaloJets*). The granularity of the HCAL and ECAL in the barrel region of CMS ($|\eta| < 1.4$) is such that one single HCAL cell is matched to a 5×5 crystal cell in the ECAL, forming a projective tower. In the endcap region, a more complex tower definition is applied. Calorimeter jets are reconstructed using energy deposits in calorimeter towers as inputs.

The anti-k_T Algorithm

The aim of jet reconstruction algorithms is to provide an estimate of the jet energy and direction that is as close as possible to the ones of the particle initiating the shower. Therefore, a successful jet algorithm should fulfill two main requirements:

- collinear-safety, meaning that the outcome does not change if the energy carried by a single particle is instead distributed between two collinear particles
- infrared-safety, that is the stability of the result of the jet finding against the addition of soft particles.

The most widely used technique for jet clustering in CMS is the anti-k_T algorithm [62], that belongs to a more general class of sequential recombination algorithms. In general, they introduce two parameters:

- d_{ij} between the two objects i, j to be clustered, i.e. the energy deposits in the calorimeter crystals
- d_{iB} between the object i and the beam.

For each pair of objects, the latter quantities are evaluated. If d_{ij} is found to be smaller than d_{iB}, the objects are combined into a single object, while if $d_{iB} < d_{ij}$ the object i is identified as a jet and removed from the list of objects to be clustered. The procedure is repeated until no entries are left. The two distance parameter mentioned above are defined as:

$$\begin{aligned} d_{ij} &= \min(k_{T,i}^{2p}, k2p_{T,j}) \frac{\Delta_{ij}^2}{R^2} \\ d_{iB} &= k_{T,i}^{2p} \end{aligned} \tag{4.1}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and $k_{T,i}, y_i$ and ϕ_i are the transverse momentum, rapidity and azimuthal coordinate of object i. The parameter R defines the maximum (η, ϕ) radius of a jet cone. According to the value chosen for the parameter p, one can have:

- the k_t algorithm, for p = 1
- the Cambridge/Aachen algorithm, for p = 0
- the anti-k_T algorithm, for p = -1.

The functionality of the anti-k_T algorithm lies in the fact that soft, low-p_T objects tend to cluster with hard ones long before they cluster among each other. When a high-p_T object has no neighbours in a radius R, then it starts to collect soft objects around it, and the result is a conical jet, provided that there are no overlaps among jets.

4.6 Particle Flow Algorithm

The particle-flow algorithm has been implemented to identify all stable particles produced in the event. For this purpose, the information coming from all CMS sub-detectors needs to be combined to obtain an optimal determination of direction, energy and type of the individual particles. The elements used by the particle-flow event reconstruction are tracks of charged particles, calorimeter clusters and muon tracks, whose reconstruction has been described in the previous Sections.

Link Algorithm A given particle is, in general, expected to be related to several particle-flow elements, that have to be connected to each other by a link algorithm; the linked elements will then represent the full reconstruction of a single particle through the detector. Starting from a pair of elements in the event, the link algorithm defines a distance between them and evaluates the quality of the link itself. Linked elements are then grouped in blocks. The link algorithm is tentatively performed on each pair of elements in the event and should avoid any possible double counting. Thanks to the high granularity of the CMS detectors, blocks typically contain only a few elements, and are used as inputs for the particle reconstruction and identification algorithm.

Particle Identification The reconstruction and identification of a set of particles from each block of elements is the core part of the particle-flow algorithm. It provides a list of reconstructed particles to be used in the subsequent physics analysis. For each block provided by the link algorithm, particle reconstruction and identification is performed according to the following steps:

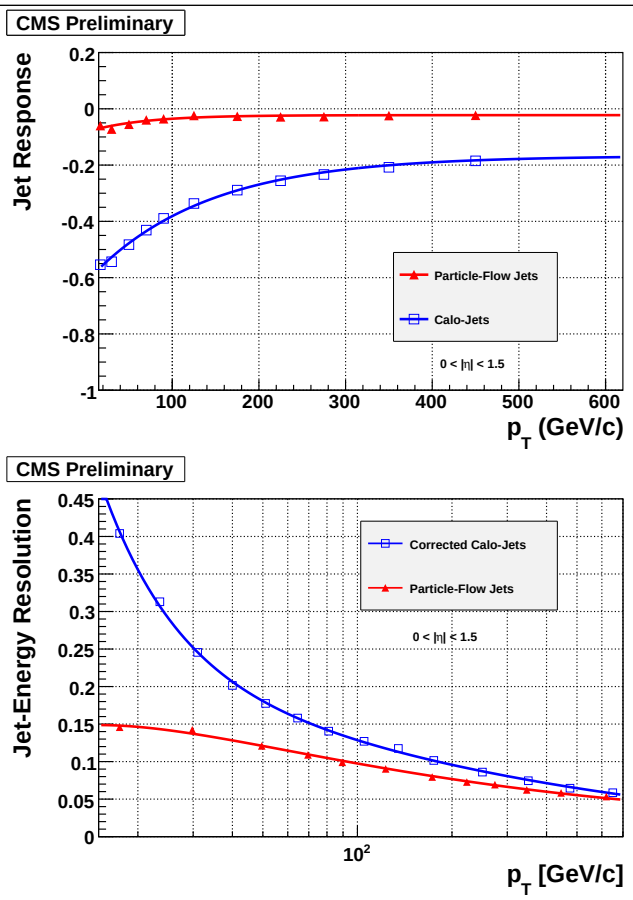
- first, each global muon whose momentum is compatible with the measurement provided by the tracker only is identified as a particle-flow muon and the corresponding track is removed from the block.
- Then, electron are identified, taking into account that their tracks tend to be short and that they lose energy by bremsstrahlung in the tracker layers. Tracks trajectories are followed all the way to the ECAL and matched with the corresponding clusters in the calorimeter. Once an electron has been identified, the associated track and ECAL clusters are removed from further processing.
- Remaining tracks in the block are associated to a particle-flow charged hadron, whose momentum and energy are taken from the tracker, assuming the pion mass hypothesis.
- The detection of neutral particles in the block, e.g. photons and neutral hadrons, is done through a comparison between the momentum of the tracks and the corrected energy detected in the calorimeters: if the energy of the closest ECAL and HCAL clusters linked to the tracks are significantly larger than what expected from the total momentum of the associated charged particles, the excess is interpreted as a photon and/or a neutral hadron.
- Finally, the remaining ECAL and HCAL clusters not linked to any track are interpreted as photons and neutral hadrons, respectively.

Performance with Jets The advantage of an event reconstruction based on the particle flow algorithm is relevant especially for jet and for missing energy reconstruction, that will be introduced in the next Section. Concerning jets, the typical energy fractions carried by charged particles, photons and neutral hadrons are 65%, 25% and 10% respectively, meaning that 90% of the jet energy can be measured with good precision by a particle-flow based jet reconstruction, only the remaining 10% of the energy being affected by the lower resolution of the hadron calorimeter. Particle-flow jets (*PFJets*) are obtained from all particles reconstructed with the particle flow, clustered according to a given algorithm for jet clustering.

Figure 4.4 compares the jet response and jet energy resolution in the barrel region, for jets built from calorimeter clusters only (CaloJets) and from particle flow candidates (PFJets), in a simulated di-jet sample. For PFJets, between 95% and 97% of the jet energy is reconstructed, depending on the jet p_T, compared to 60%-80% for CaloJets. In addition, the gain in resolution for PFJets can reach up to a factor of 3 with respect to CaloJets.

Figure 4.5 shows the improvement in angular resolution that can be achieved using PFJets rather than CaloJets in the barrel region, again evaluated on a di-jet simulated sample.

Figure 4.4 Jet response (top) and jet energy resolution (bottom) for simulated di-jets events in the barrel region. Jets reconstructed with particle flow candidates (red triangles) or with the calorimeters clusters only (blue open squares) are compared [63].

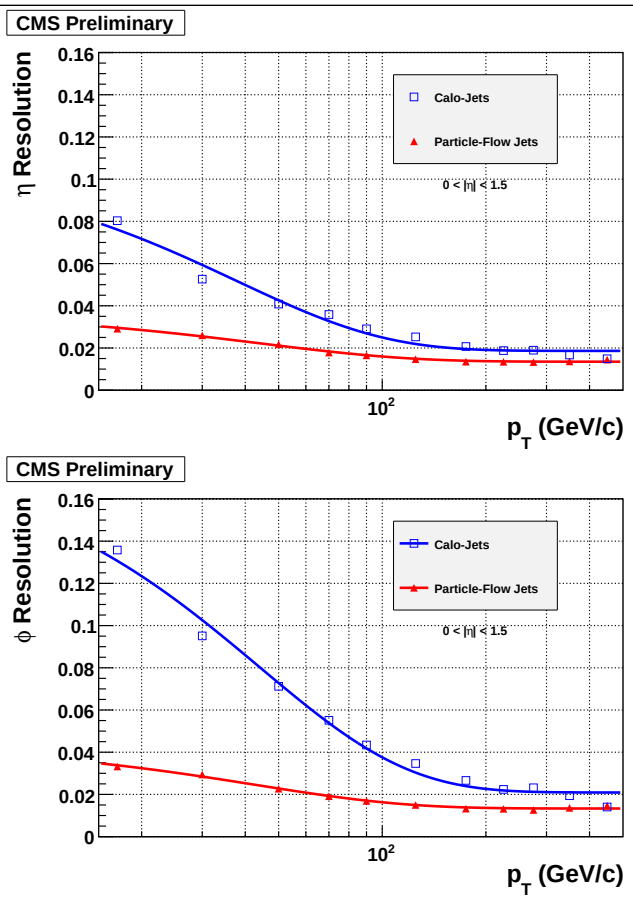


4.7 Missing Transverse Energy

The CMS detector can efficiently detect most variety of particles produced in proton collisions, with the exception of neutrinos and hypothetical neutral weakly interacting particles that have not been discovered yet. Due to their low interaction cross section, these particles do not produce any signal in the detector. However, their presence can be inferred from the momentum imbalance in the plane perpendicular to the beam direction. Indeed, assuming that the incoming protons carry no transverse momentum, momentum conservation applied on the transverse plane implies that the vectorial sum of transverse momenta of all final state particles must be zero as well. When it is not the case and there is missing transverse momentum in order to satisfy momentum conservation, it is the hint for the presence of invisible particles, e.g. neutrinos. Missing transverse momentum (or energy) is usually referred to as MET.

MET reconstruction relies on precise calibration for all physics objects and it is particularly sensitive to detector malfunctions and pileup interactions. Several algorithms for missing energy reconstruction have been developed in CMS [64]. Although MET

Figure 4.5 Jet η (top) and ϕ (bottom) resolutions for simulated dijet events in the barrel region. Jets reconstructed with particle flow candidates (red triangles) or with the calorimeters clusters only (blue open squares) are compared [63].



reconstruction can in principle be performed with the only information of energy deposits in the calorimeters, most physics analysis prefer to use MET reconstructed using a particle-flow tecnique, as it gives a better performance.

The magnitude of the MET can be underestimated for many reasons, e.g. minimum energy thresholds in the calorimeters, p_T thresholds and inefficiencies in the tracker, and the nonlinearity of the calorimeter response for hadrons. This bias is significantly reduced if the MET is calculated after correcting the p_T of the jets using jet energy corrections. Further corrections have been introduced to improve the performance of missing energy reconstruction in events with a large number of pileup interactions.

MET can also be over estimated: anomalous high-MET events can originate from several sources, e.g. particles striking sensors in the ECAL barrel detector or dead calorimeter cells and electronic noise. Such events have been studied in detail [65] and cleaning algorithms have been developed and applied. Figure 4.6 shows the MET distribution in dijet events before and after the cleaning algorithms have been applied.

The performance of MET reconstruction has been estimated from data using $Z \to \mu\mu$, $Z \to ee$ and photon events. It is done by removing the well reconstructed vector boson

(or photon) from the list of reconstructed particles and measuring the MET from the recoil of the hadronic system. Figure 4.7 shows the resolution of the recoil component parallel (left) and perpendicular (right) to the direction of the Z boson (or γ) transverse momentum.

Figure 4.6 Particle Flow MET distribution for events passing a dijet selection, before and after applying cleaning algorithms to remove anomalous high-MET events [64].

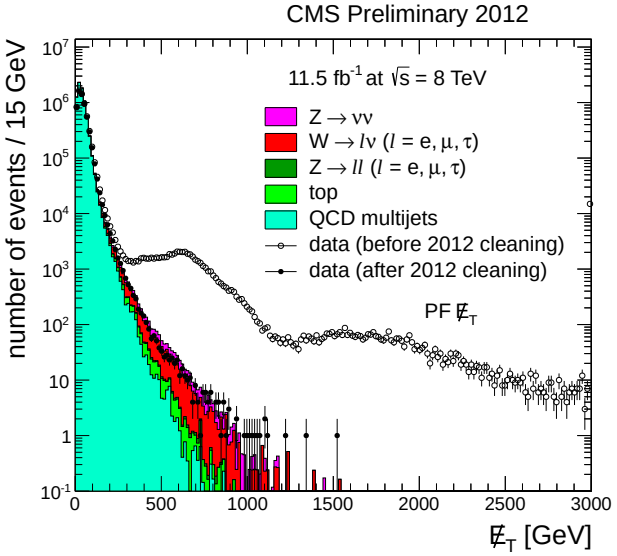
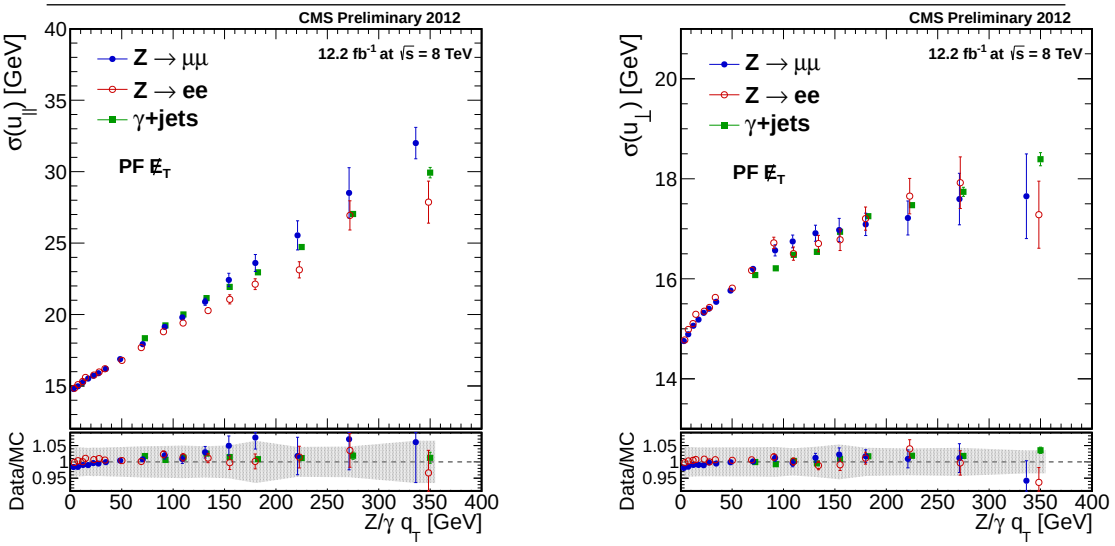


Figure 4.7 Resolution curves of the parallel recoil component $u_{\parallel}$ (left) and perpendicular recoil component $u_{\perp}$ (right) as a function of the Z/γ transverse momentum q_T for PF MET. The upper frame shows the response in data; the lower frame shows the ratio of data to simulation, the grey band indicating the uncertainty [64].

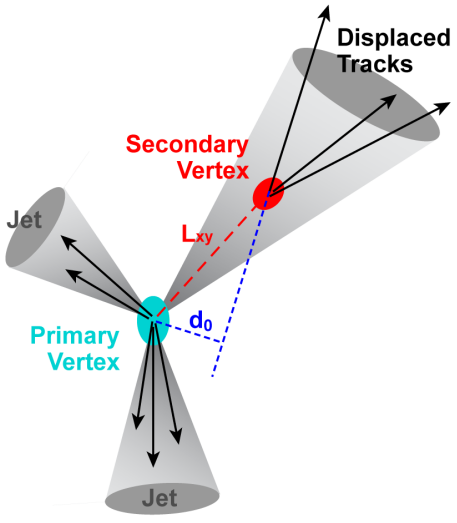


4.8 Identification of b-jets

The identification of jets from the hadronization of bottom quarks, usually referred to as *b-jets*, is a crucial element for many physics analysis. Some particular features of B hadrons and of their decays makes it possible to distinguish whether a jet is more likely to originate from a b-quark rather than from another flavour quark (or from a gluon). The following paragraphs illustrate the main concept of b-tagging and the algorithms developed by the CMS collaboration.

Bottom quarks are the heaviest quarks that hadronize and therefore show up in the detector as jets, as the heavier top quarks have a too short lifetime ($\tau_{top} \sim 3 \cdot 10^{-25}$ s) and decay before creating any hadronic bound state [66]. Hadrons containing b flavoured quarks, e.g. B mesons or Λ baryons, are characterized by their relatively large masses, sizable lifetimes and hard momentum spectra of daughter particles. Moreover, b-jets often ($\sim 20\%$) contain a muon from semileptonic decays of B hadrons. The relatively long lifetime of B hadrons ($c\tau \sim 500$ μm) is such that they travel a sizable distance before decaying in the tracker volume. Therefore, tracks from charged daughter particles identify a displaced secondary vertex within the jet.

Figure 4.8 A simplified illustration of the b-jet identification concept [67]. B hadrons decays leads to the presence of displaced tracks with large impact parameters with respect to the primary vertex, and they possibly identify a secondary vertex in the jet, due to the sizable lifetime of B hadrons.



When a secondary vertex is reconstructed within a jet, the flight distance between the primary vertex of the interaction and the secondary is a useful physical observable to be exploited for b-tagging. Tracks originating from decay of the hadron do not point to the primary interaction vertex. The impact parameters of such tracks, defined as the minimum distance between each track and the primary vertex, are therefore expected to be larger than the ones of tracks arising from a primary interaction. Figure 4.8 is a simplified illustration of a b-jet. A good ability in b-jet identification especially relies on

efficient and high quality track reconstruction, and also on an efficient muon detection. The CMS detector with its silicon tracker comprising three pixel layers and an extended muon system matches well the b-tagging requirements.

B-tagging Algorithms

The final goal of a b-tagging algorithm is to provide a discriminant on a jet by jet basis, whose value gives an indication of how likely it is for the jet to originate from a bottom (rather than another flavour) quark. All the algorithms described in the following paragraphs start from the collection of tracks associated to each jet and discard the ones that do not satisfy quality requirements, e.g. with an insufficient number of hits in the pixel detector or with a high value of the χ^2 of the fit. Three categories of b-tagging algorithms can be identified:

- track based algorithms discriminate using information from single tracks in the jet
- vertex based algorithms require the presence of a reconstructed secondary vertex in the jet cone and discriminate using its properties
- combined algorithms combining the two previous approaches.

Track Based Algorithms

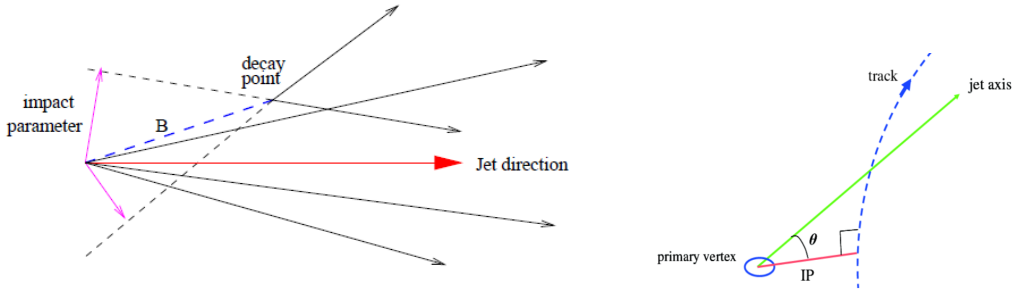
Tracks from in flight decay of B hadrons are expected to be displaced from the primary vertex and therefore to have a relatively large impact parameter (IP), defined as the distance between the primary vertex and the point of closest approach of the track. Thanks to the high resolution on the measurement of the z coordinate provided by the pixel tracker, the IP is calculated in three dimensions. Despite being a distance, a sign can be assigned to the IP, defined as the scalar product of the vector pointing from the primary vertex to the point of closest approach of the track with the jet direction. Therefore, tracks originating from particles decaying while traveling along the jet axis tend to have positive IP values, while prompt tracks can have either positive or negative impact parameters. Figure 4.9 illustrates the impact parameter definition and how its sign is defined. The resolution on the impact parameter depends on many factors, such as the tracker geometry and single hit resolution, as well as the momentum and direction of the track. The precision in measuring the IP can be very different from track to track and it is of the same order of magnitude of the IP itself. For this reason, the observable used is the impact parameter significance S_{IP}, defined as the ratio of the IP divided by its uncertainty.

The impact parameter significance has itself discriminating power between b- and non-b jets, and it is used as discriminator value for the Track Counting (TC) algorithms. For each jet, it sorts the associated tracks in decreasing order of IP significance, and the value of the second (or third) track is used as a discriminator for the Track Counting High

Efficiency (or High Purity) algorithms. The distribution of the TCHE discriminator is shown in Figure 4.10.

An extension of the TC is the Jet Probability algorithm, that exploits the IP significance information of all tracks in the jet and a discriminator is built estimating the likelihood that all tracks in the jet come from the primary vertex. The distribution of the JP discriminator is provided in Figure 4.10. A slightly different version is the Jet B Probability (JBP) algorithm, that gives more weight to the tracks with the highest IP significance, up to four such tracks, as this is the average number of reconstructed tracks from a B hadron decay.

Figure 4.9 An illustration of the impact parameter definition (left). The sign of the impact parameter is assigned accordingly to the cosine of the angle between the IP vector and the jet axis direction (right) [68].



Secondary Vertex Based Algorithms

The presence of a reconstructed secondary vertex (SV) within a jet provides another handle to identify b-jets by taking into account the flight distance and direction of the SV and several kinematic variables of the associated tracks, such as their multiplicity, energy and mass (assuming all tracks to come from pions). The left plot of Figure 4.11 shows the number of reconstructed secondary vertices within the jet cone, in a multijet sample. To enhance the purity and reduce fakes, some quality requirements are applied to secondary vertices candidates, e.g. on the fraction of tracks shared with the primary vertex. Additional requirements are applied to reduce contamination from interaction of particles with the detector material and by decays of long lived mesons. The Simple Secondary Vertex (SSV) algorithms discriminate on the significance of the three dimensional flight distance between the primary and the secondary vertices. The right plot of Figure 4.11 shows the distribution of the flight distance significance for multijet events. As for the TC algorithm, two versions are available. The High Efficiency SSV (SSVHE) allows secondary vertex reconstruction with at least two tracks, while the High Purity (SSVHP) version asks for a minimum of three tracks.

Figure 4.10 Distributions of the TCHE discriminator (top) and of the JP discriminator (bottom) in multijet events, in simulation and in 7 TeV data. [69].

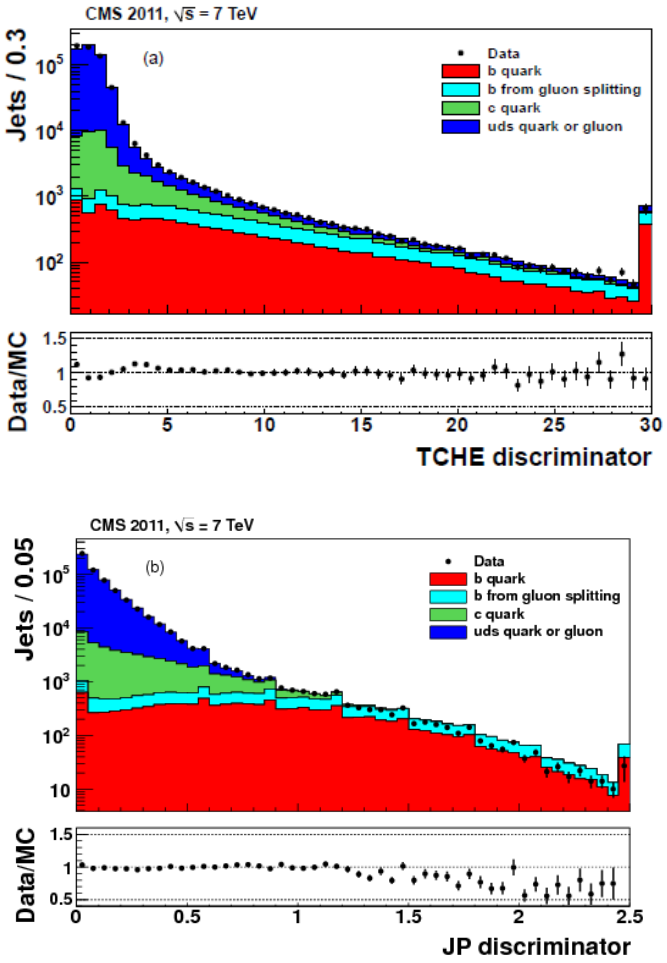
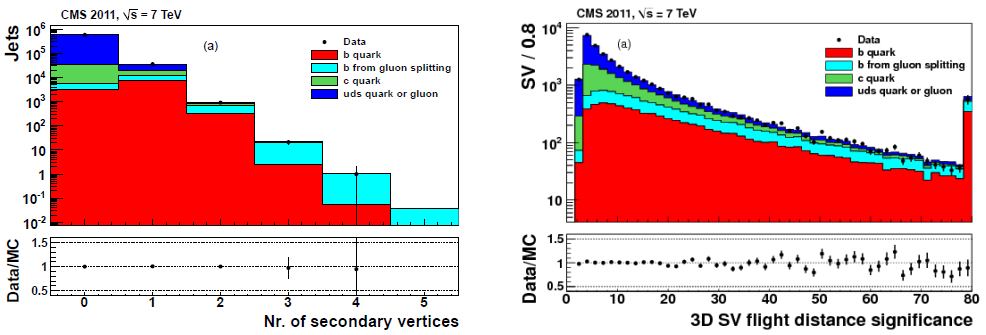


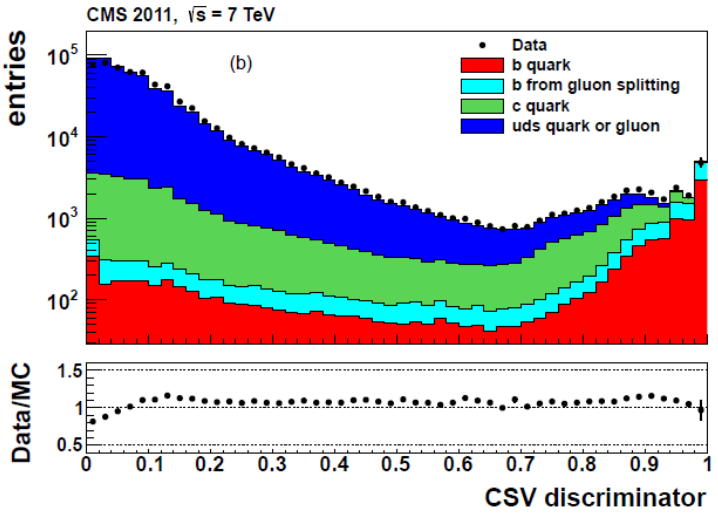
Figure 4.11 Number of reconstructed secondary vertices inside the jet cone (left) and the distribution of the three dimensional distance between the primary and the secondary vertex, in multijet events from 7 TeV data and in simulations [69].



Combined Taggers

All secondary vertex based algorithms are limited by the ability of identifying secondary vertices, whose reconstruction efficiency is $\sim 60 - 70\%$. Efficiency can be recovered when using a combined approach of track and SV based algorithms: when no SV is reconstructed, the jet tagging can still be done using tracks information. The Combined Secondary Vertex (CSV) algorithm combines information from several variables, like the multiplicity of tracks in the jet, the flight distance, the secondary vertex mass, the energy carried by tracks with respect to the total jet energy and the IP significance of the first track with a mass higher than the charm quark mass. Two likelihood ratios are built and used to discriminate between b and light parton jets and between b and c jets. The distribution of the CSV discriminator is presented in Figure 4.12. The CSV output covers the range $(0, 1)$. Jets with a CSV value close to 1 are most likely from b-quarks, while light jets populate the lower part of the CSV distribution.

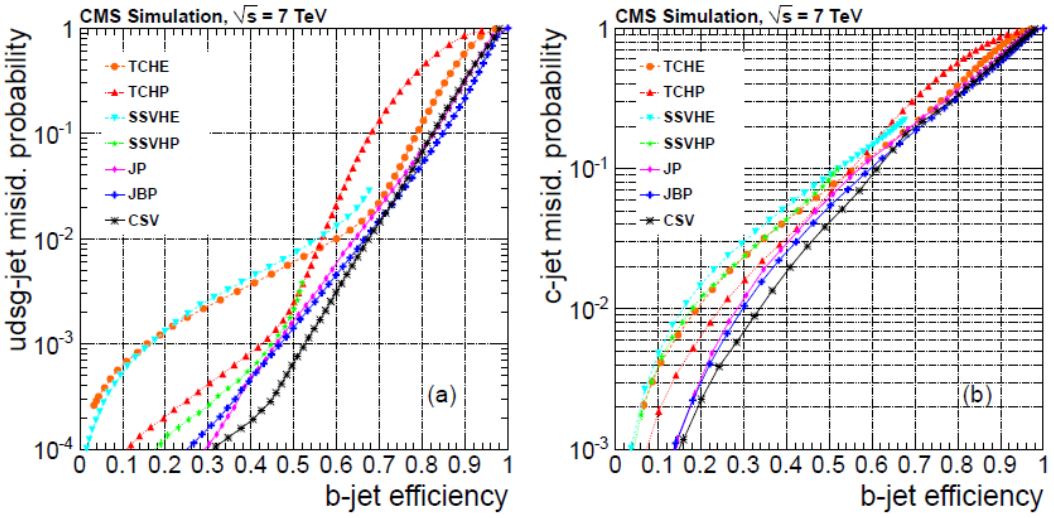
Figure 4.12 The distribution of the CVS discriminator, in a multijet simulated sample and in 7 TeV data [69].



Efficiency and Mistag Rate The performance of a b-tagging algorithm can be estimated through its ROC (Receiver Operating Characteristic) curve, that represents the fake rate as a function of the b-jet efficiency. The fake rate is the probability of addressing a jet as coming from a b-quark, when it is actually originating from another flavour quark, while the b-jet efficiency is the probability of correct assignment of b-jets to b-quarks. Fake rates and the b-jet efficiencies can be easily measured from Monte Carlo samples, as well as from data, using more sophisticated techniques. Different algorithms can be compared through their ROC curves, the best algorithm being the one that provides the lowest mistag rate for a given b-jet efficiency. Due to physical reasons, the probability of identifying a jet from a charm quark as a b-jet is much higher than the mistag rate for the other light flavour quarks (u,d,c,s) or gluons, because charmed

hadrons are more similar to B hadrons in terms of masses and lifetimes. For this reason, separate ROC curves are drawn for c-jet and light flavour (udscg) jets mistag rates. Figure 4.13 shows the ROC curves for some b-tagging algorithms used in Run 1, evaluated from multijet simulated simulated events, after applying a p_T threshold of 30 GeV. The CSV algorithm performs best for the b-efficiency working points chosen by most of the analyses, and was therefore the most commonly used.

Figure 4.13 ROC curves for several b-tagging algorithms used by CMS in Run 1 analysis. Mistag rates for light udsg jets (left) and for c-jets (right) are shown as a function of the b-jet efficiency in multijet simulated events [69].



Chapter 5

Search for VH Production in the $b\bar{b}$ Decay Channel from Run 1 to Run 2

In the first Section of this Chapter the strategy and results for the search of the Higgs boson produced in association with a vector boson and decaying to a $b\bar{b}$ pair are presented. The second Section describes the preparation of the analysis for the upcoming LHC run at a center of mass energy of 13 TeV. First, a comparison between Monte Carlo samples at 8 and 13 TeV at generator level, to understand the difference in event kinematics is performed. Then, the performance of a simple cut-based invariant mass analysis is evaluated using samples of signal and main background events that have been processed through a full CMS detector simulation. The goal is to estimate the sensitivity of a cut-based $Z(ll)H(b\bar{b})$ search in the electron and muon channels and compare it to the 8 TeV results.

5.1 VH, H(bb) Search with Run 1 LHC Data

At the LHC the search for the Higgs boson in associated production with a vector boson and decaying to a bottom quark pair at the LHC has been performed by the CMS [70] and ATLAS [71] experiments, using data collected between 2009 and 2012. In this period, usually referred to as the Run 1 of the LHC, protons were first collided at a center of mass energy of 7 TeV, then the energy was increased to 8 TeV. In this Section the main points of the strategy chosen by the CMS collaboration for the VH,H(bb) analysis are presented. The integrated luminosities recorded by the CMS detector at 7 and 8 TeV were 5.1 and 19.7 fb^{-1}, respectively.

5.1.1 Analysis Strategy

The search for VH associated production in the Higgs to $b\bar{b}$ decay channel has been performed considering several final states, that differs from the vector boson being either a W or a Z and from its decay mode. The following channels have been considered:

- Z(ll)H(bb), with $l = e, \mu$
- Z($\nu\nu$)H(bb)
- W($l\nu$)H(bb), with $l = e, \mu$
- W($\tau_{had}\nu$)H(bb), where the τ lepton decays hadronically.

Boost Regions The analysis is performed in a boosted regime, meaning that events are selected where the Higgs and the vector bosons have large transverse momentum. This choice is advantageous for two reasons. First, the cross section for the production of a vector boson in association with jets, that is the main background, decreases more rapidly than the signal cross section as a function of the vector boson transverse momentum, therefore the signal over background ratio increases in the boosted regime. Then, in such boosted regime the Higgs and the vector boson recoil against each other and the consequent back-to-back topology of signal events provides an additional handle to reject background events.

Trigger Events have been recorded using several trigger paths, mainly requiring the presence of one or two electrons, one muon or missing energy with a transverse momentum above a given threshold, depending on the signal final state. For the single lepton triggers, the p_T threshold ranged from 17 to 24 GeV for muons and from 17 to 30 GeV for electrons, depending on the LHC running conditions. Dedicated triggers exploiting b-tagging or jet plus MET kinematics have been used as well for the W($l\nu$)H and Z($\nu\nu$)H channels.

Physics Objects Particle candidates are reconstructed from particle-flow objetcs. Electrons and muons are required to be central ($|\eta| < 2.4$) and isolated to reject candidates produced from hadron decays inside a jet. Moreover, electrons are required to have $p_T > 30$ and 20 GeV in the W($e\nu$)H and Z(ee)H analyses, respectively. A similar p_T threshold of 20 GeV is applied for muons. Taus are reconstructed using the Hadron Plus Strip algorithm [72] in the region $|\eta| < 2.1$ and requiring $p_T > 40$ GeV. Additional quality cuts are applied to reduce the fake rate due to muons and electrons.

Jets are built by clustering particle-flow objects according to the anti-k_T algorithm with a radius parameter $R = 0.5$. They are required to lie in the region $|\eta| < 2.5$, to have at least two associated tracks and their electromagnetic and hadronic energy fractions to be at least 1% of the total energy. Finally, a p_T cut with a threshold of 30 GeV

is applied for all channels, with the exception of the cleaner dilepton Z(ll)H channel, where a lower cut at $p_T > 20$ GeV is found to be optimal. Identification of jets coming from the hadronization of a b-quark, usually referred to as b-tagging, is possible thanks to some peculiar features of b-hadrons, i.e. their sizeable lifetime. Several algorithms have been developed in CMS to serve this purpouse. This analysis uses the Combined Secondary Vertex (CSV) algorithm for b-jets identification, that has been presented in Chapter 4.

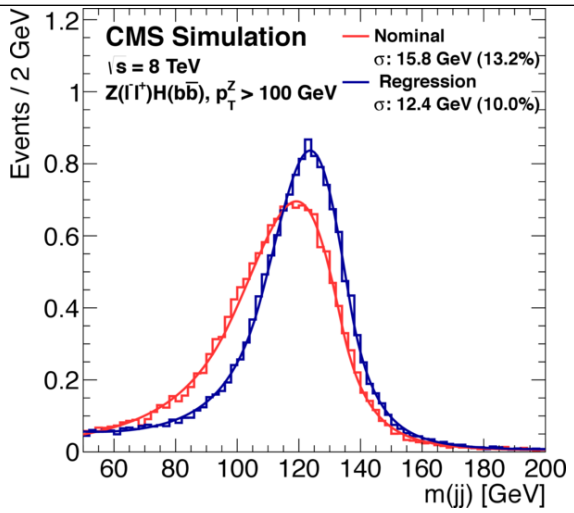
Missing transverse energy is measured as the transverse momentum needed to balance the sum of momenta of particle-flow objects on the transverse plane. For Z($\nu\nu$)H events the MET is required to be greater than 100 GeV.

Bosons Reconstruction Z boson candidates in the dilepton channel are reconstructed combining same flavour, opposite charge leptons. To be compatible with a Z decay, the dilepton invariant mass is required to lie within the interval $[75 - 105]$ GeV. Selection of events in the Z($\nu\nu$) channel is done just requiring the MET to be above a given threshold. In this channel, the MET is a measurement of the Z candidate p_T. For the WH channels, event candidates are identified from the presence of a single isolated lepton plus missing transverse energy, from which the vector boson can be reconstructed in the transverse plane.

The vector boson p_T distribution in data is found to be softer than in the simulation, such difference is expected to arise from higher order electroweak corrections to the simulation [73][74]. To correct for this effect, a reweighting of the vector boson p_T spectrum is derived by fitting the data/MC ratio in a region dominated by V+light jets events, separately for each channel, and applied in the signal region. Moreover, calculated estimates of the NLO electroweak and NNLO QCD corrections to VH production in the boosted regime have been applied to Monte Carlo events.

The reconstruction of the Higgs boson decaying to a b-quark pair is challenging due to the hadronic final state. This step is performed by identifying the pair of jets with the largest total transverse momentum $p_T(jj)$. The event is discarded if the two selected jets do not satisfy some minimum b-tagging requirements. In order to improve the resolution on the dijet mass, a b-jet energy regression is also applied. With such technique the energy of b-jets assigned to the Higgs decay is corrected in the attempt to remove the bias and resolution smearing due to the presence of neutrinos from semileptonic decays of B hadrons. The correction factor is computed individually for each jet, using a multivariate analysis (MVA) technique that takes into account several variables, e.g. energy, p_T and number of constituents of the jet, as well as information about leptons inside the jet itself. The regression is trained on simulated signal events and validated on a Z+bb data sample. After the regression is applied, the resolution on the invariant mass of the two b-jets $m(bb)$ improves of ~ 10 to 30% and the mass peak is found to be closer to the nominal Higgs mass, as can be seen in Figure 5.1. The regression has also been validated on $t\bar{t}$ events, showing improvements on the top mass resolution.

Figure 5.1 Dijet invariant mass distribution for simulated samples of Z(ll)H(bb) events ($m_H = 125$ GeV), before (red) and after (blue) the b-jet energy correction from the regression procedure is applied [70].



Signal and Background Topologies Signal events are characterized by the the presence of a vector boson recoiling against the Higgs boson candidate, reconstructed from two b-jets with an expected invariant mass around 125 GeV. The transverse momentum of the b-jets peaks around 50 GeV and the dijet p_T for signal events is relatively harder than in background events. No isolated leptons are expected in addition to the ones from W or Z bosons decays. In the boosted regime chosen for the analysis, the vector and Higgs bosons candidates are expected to be central and back-to-back in the transverse plane.

The main backgrounds for this analysis are:

- the production of W and Z in association with jets (V+jets). It has a final state very similar to the signal, however the p_T spectrum is softer on average and the jet invariant mass distribution does not show any resonance peak in the region around the expected Higgs mass. The largest contribution after applying b-tagging requirements arises from V+bb events.

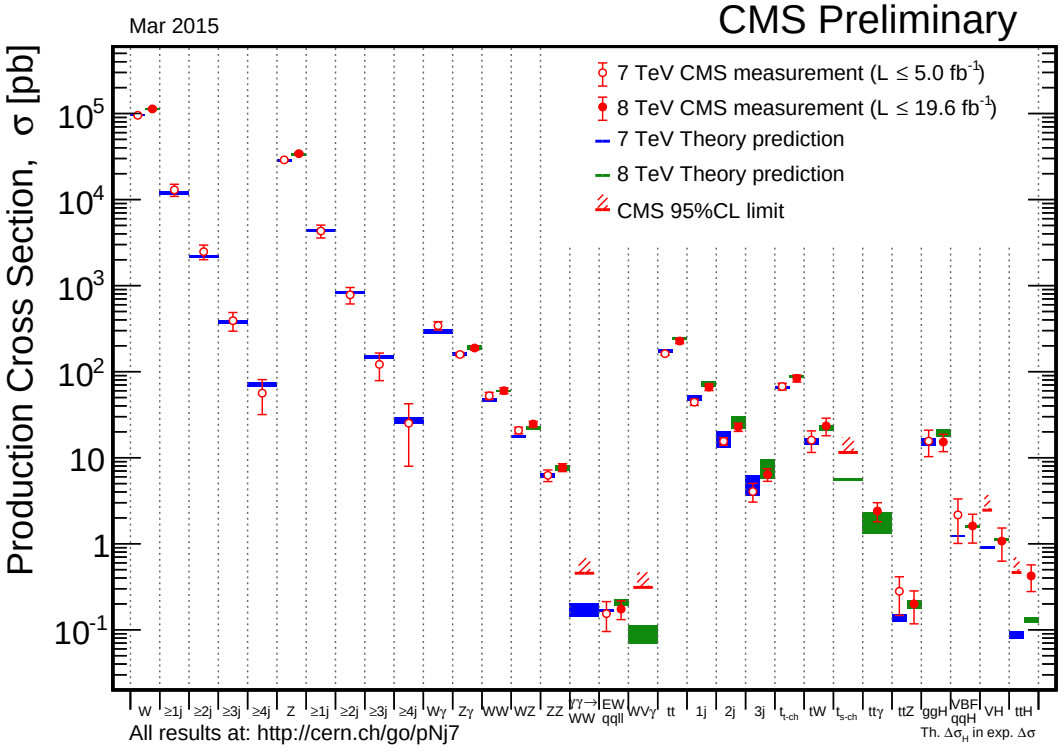
- Top quark production, both single top and $t\bar{t}$ pairs. It has a relatively high cross section at the LHC, and a final state that include one or two real W decays, possibly to leptons, and b-jets, as well as missing energy from the W decays. Single top events are mostly important as a background for the WH channels, while the Z(ll)H channel is less affected. This background source can be distinguished from the signal because the average jet multiplicity in top events extends well beyond two, additional leptons of different flavours can be present and the azimuthal opening angle between the vector boson and dijet is more broadly distributed than in signal events.

- The production of vector boson pairs (WW,WZ,ZZ). It is an almost irreducible background, as the vector boson pair can decay to a jet pair having an invariant

mass close to what expected from signal events. The most difficult final state to be rejected is when one boson decays leptonically and the second one hadronically to a jet pair. The use of b-tagging does not help in rejecting the background when a Z boson decays to a bottom quark pair, and the only handle in this case is provided by the dijet mass.

Figure 5.2 shows the inclusive production cross sections for several processes, including the above mentioned backgrounds, measured by CMS in proton collisions at a center of mass energy of 7 and 8 TeV.

Figure 5.2 Inclusive production cross sections for several processes, measured in proton collisons at at $\sqrt{s} = 7$ and 8 TeV. The most important background process for the VH search are the production of W and Z in association with jets, top quark pair and single top production and diboson production [75].



Multivariate Analysis The analysis is based on a multi-variate approach allowing to extract more information from the events by exploiting the correlation of variables. A Boosted Decision Tree (BDT) [76] class is implemented using the TMVA [77] toolkit. Boosted Decision Trees are trained on Monte Carlo events to discriminate between signal and background, basing on several kinematic variables available. The training procedure ranks the variables according to their discriminating power and finds selection criteria that give the best separation between signal and background events. The training of the BDTs is performed separately for each channel and for each mass hypothesis for the Higgs boson mass.

Several variables are used as inputs to the trees:

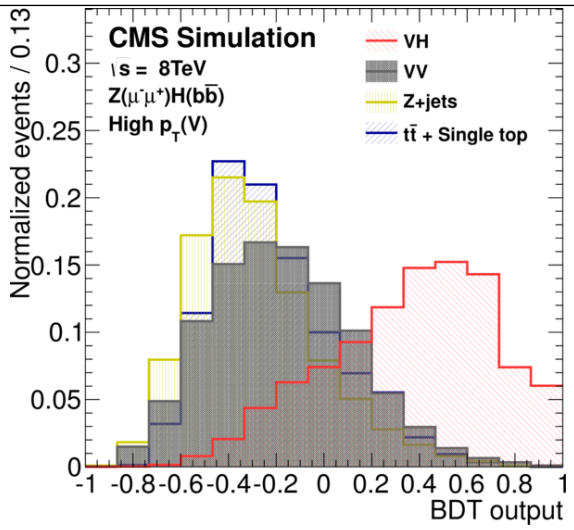
- m(jj) the invariant mass of the two jets used to reconstruct the Higgs candidate, namely the Higgs boson mass
- $p_T(j)$ the transverse momentum of the Higgs daughters
- $p_T(V)$ the transverse momentum of the vector boson candidate
- CSV values of the Higgs jets
- $\Delta\phi$(V,H) the azimuthal angle between the Higgs and the vector bosons
- $\Delta\eta(jj)$ the difference in η between the two Higgs jets
- $\Delta R(jj)$ the η-ϕ distance between the Higgs daughters
- N_{aj} the number of additional jets in the event, lying the central η region and above a given p_T threshold
- other angular variables, e.g. $\Delta\phi$ between the MET and the closest jet for the Z($\nu\nu$) channel.

Once the BDT is trained, it is used to analyze events: for each input event it returns an output, typically a number between -1 and 1, depending on how likely it is to be a signal event. Events to be processed through the BDT are required to pass preselection cuts, that are quite loose because the BDT is already meant to discriminate between signal and background.

Figure 5.3 shows an example of BDT output distribution for simulated events in the Z($\mu\mu$)H channel, in the high boost region. It can be seen how signal and background events have different distributions, the former peaking at high values and the latter being shifted towards lower ones. In order to obtain the number of signal and background events from data, a likelihood fit is performed to the distribution of the BDT output, using as templates the distributions for signal and background obtained from simulated events.

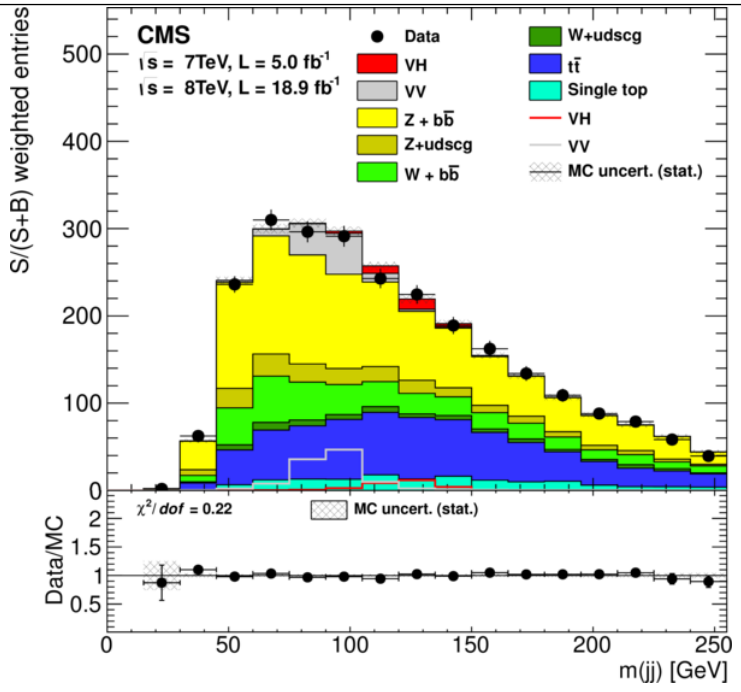
To validate the MVA approach and demonstrate that it is actually capable of identifying a small signal over a large background, the same analysis strategy has been applied to search for diboson production. Diboson production of WZ and ZZ, with a Z decaying to a bottom quark pair, has in fact a cross section that is only seven times higher than the VH,H(bb) cross section, and an almost identical final state. Moreover, cross sections for such processes are predicted by the Standard Model, allowing to compare the result obtained from the MVA analysis with the predictions. Using the same strategy as for the VH search, the diboson signal strength has been measured to be $\mu_{VV} = 1.19 \pm 0.25$, the observed and expected significance being 7.5 σ and 6.3 σ respectively.

Figure 5.3 BDT output distributions, normalized to unity, for Z($\mu\mu$)H simulated events in the high-boost region, for all backgrounds and signal [70].



Invariant Mass Analysis An invariant mass analysis has also been performed. However, due to its lower sensitivity, it is meant only to crosscheck the results obtained with the MVA approach. In this case, the background rejection relies on cuts on the kinematic variables, that are optimized for each channel and p_T bin. The number of signal and background events are then obtained performing a fit to the shape of the Higgs mass distribution. Figure 5.4 shows the invariant mass distribution of the dijet system, weighted and combined over all final states, after applying the cuts defining the signal region.

Figure 5.4 Weighted dijet invariant mass distribution, combined for all channels, in 7 and 8 TeV data and in simulations [70].



Data Driven Background Estimate The estimate of the different backgrounds contribution is based on Monte Carlo. Normalization of the main backgrounds is corrected using a scale factor obtained from data. For this purpouse, control regions enriched in a given background are identified by reverting or removing some cuts. Three control regions has been defined for the main backgrounds, namely $t\bar{t}$ production and production of W and Z bosons in association with light flavour and heavy flavour jets. A set of simultaneous fits is then performed to the control regions separately in each channel to obtain consistent data/MC scale factors, to be applied as a correction in the signal region.

5.1.2 Results of the VH,H(bb) Search with Run 1 Data

Figure 5.5 shows the result of the BDT-based analysis, for all channels combined and for the full Run 1 dataset. Events are ordered based on the signal over background ratio (S/B) of the corresponding BDT output bin, estimated from the simulation. An excess of events is observed in the bins with the higher S/B ratio. It is compatible with the signal yield expected from a Standard Model Higgs boson with a mass $m_H = 125$ GeV. The significance of the excess in terms of local p-value corresponds to 2.1 σ for $m_H = 125$ GeV, as can be seen in Figure 5.6.

Figure 5.5 Combination of all channels into a single distribution. Events are sorted in bins of similar expected signal-to-background ratio, as given by the value of the output of their corresponding BDT discriminant (trained with a Higgs boson mass hypothesis of 125 GeV). The two bottom insets show the ratio of the data to the background-only prediction and to the predicted sum of background and SM Higgs boson signal with a mass of 125 GeV [70].

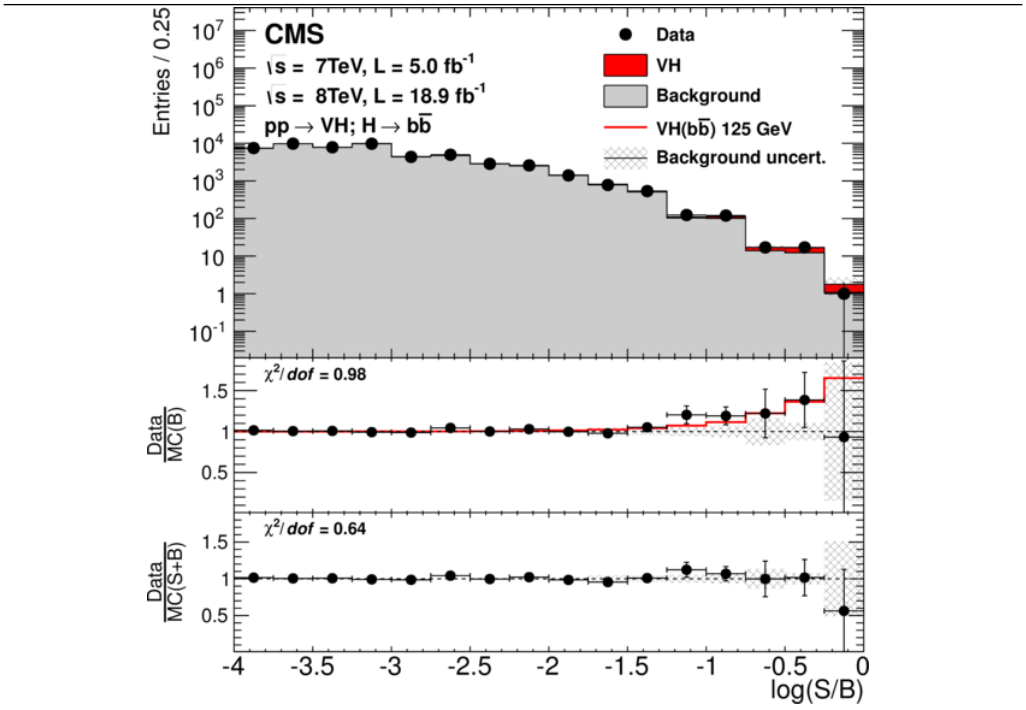
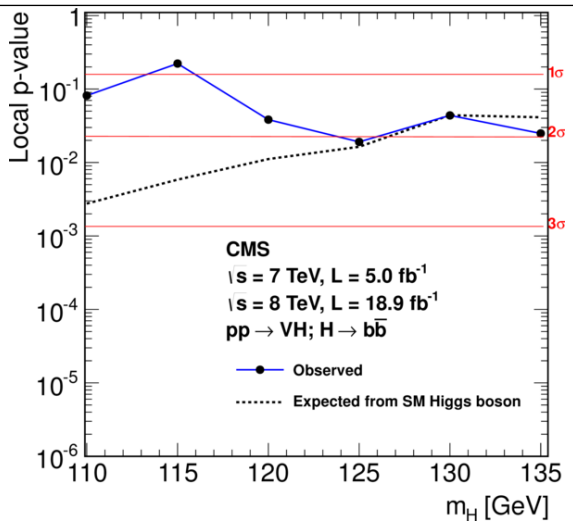
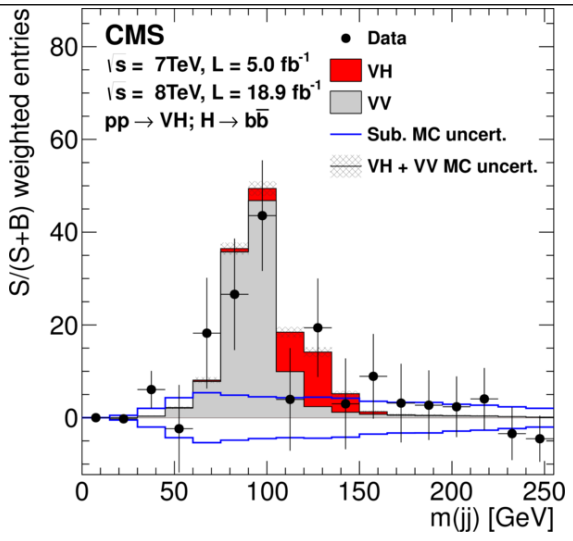


Figure 5.6 Local p-values and corresponding significance for the background-only hypothesis to account for the observed excess of events in data [70].



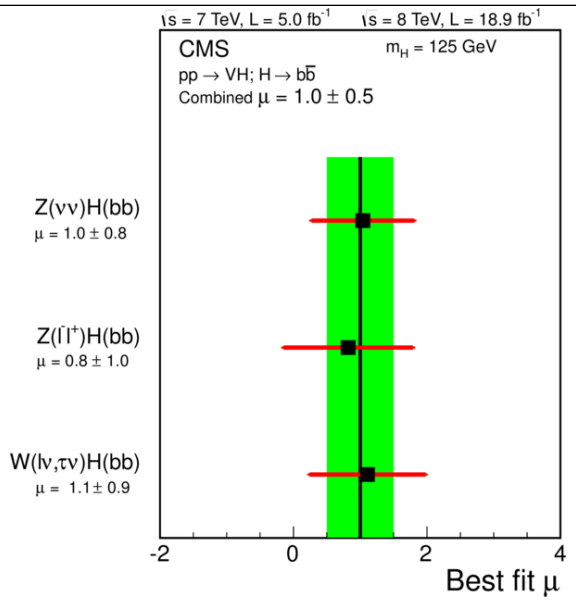
The invariant mass distribution as obtained from the crosscheck analysis after background subtraction is shown in Figure 5.7. An excess of events on top of the diboson background is observed, and is again compatible with a SM Higgs boson. For the invariant mass analysis, the measured signal strength, relative to that predicted by the Standard Model, is obtained from a fit to the m(jj) distribution and is found to be $\mu = 0.8 \pm 0.7$. The excess of events, taking into account all considered channels, has a local significance of 1.1 σ with respect to the background-only hypothesis. For the Z(ll)H(bb) channel, the observed and expected local significance amounts to 0.35 and 0.60 standard deviations, respectively. The relative contribution of statistical and systematic uncertainties can be expressed as an uncertainty on the expected value of the signal strength $\mu^{exp}_{Z(ll)} = 1.00 \pm 1.55(\text{stat}) \pm 0.47(\text{syst})$.

Figure 5.7 Weighted dijet invariant mass distribution, combined for all channels, after all backgrounds except the diboson events have been subtracted [70].



The measured best fit values for the signal strength μ are provided in Figure 5.8, separately for the WH and ZH channels and combined. The best estimate of the VH cross section times the H(bb) branching ratio, normalized to its Standard Model expectation value, is found to be $\mu_{VH} = 1.0 \pm 0.5$. The expected and observed upper limits for all considered mass hypothesis are shown in Figure 5.9.

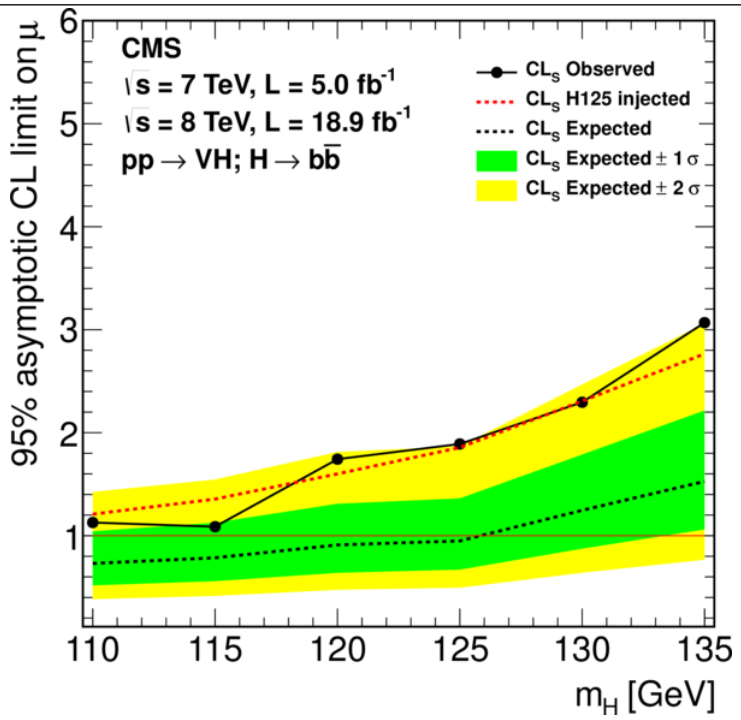
Figure 5.8 The best-fit value of the production cross section for a 125 GeV Higgs boson relative to the standard model cross section, i.e., signal strength μ, for partial combinations of channels. The green band shows the result for all channels combined [70].



Uncertainties The total uncertainty on the final results have comparable statistical and systematic contributions. The most important contributions to the systematic uncertainty arise from MC modeling and data driven estimate for the backgrounds, as well as from b-tagging efficiency. Uncertainties on the jet energy scale and resolution as well as on the signal cross section are also contributing.

Updated Results The results presented so far [70] have been updated when differential calculation of the gg→ZH process as a function of Z p_T became available, leading to an increase of the expected signal yields of about 10÷30% in the boosted region. When combining with the WH channel, the expected significance increases from 2.1 to 2.5 σ, while the observed significance remains unchanged at 2.1 σ [78]. The best fit value for the signal strength is found to be $\mu_{VH} = 0.89 \pm 0.43$ [79].

Figure 5.9 The expected and observed 95% CL upper limits on the product of the VH production cross section times the Higgs to $b\bar{b}$ branching fraction, with respect to the expectations for the SM Higgs boson. The limits are obtained combining the results of the searches using the 2011 (7 TeV) and 2012 (8 TeV) data. The red dashed line represents the expected limit obtained from the sum of expected backgrounds and the SM Higgs boson signal with a mass of 125 GeV [70].



5.2 Preparations of the Z(ll)H(bb) Analysis for Run 2

In the following paragraphs, the first studies on simulated events to prepare the Z(ll)H(bb) analysis for the 13 TeV run are presented. First, some kinematic distributions at generator level, that is without simulating the detector reconstruction, have been considered and compared to the corresponding 8 TeV samples. Then, samples of reconstructed events at 13 TeV for the Z(ll)H(bb) signal and most important backgrounds have been taken into account and used to perform a simplified invariant mass analysis. At the time of preparing this Thesis, the set of available Monte Carlo events was limited to the pre-production samples used mainly to validate the computing and reconstruction performance to prepare for the Run 2 data taking and analyses. For this reason, some samples that are in principle needed for the Z(ll)H(bb) analysis, e.g. diboson processes with two leptons and two jets final state, are missing in the following studies. While such samples are needed for the final analysis, neglecting them in the signal over background estimates reported in this Thesis have minor effect, as they represent only approximately 10% of the total background yield, even in the most diboson enriched mass bin.

5.2.1 Basic Kinematics Studies

A basic study of event properties changes in the 8 to 13 TeV energy transition has been performed using Monte Carlo events after parton showering, without any detector simulation. For these studies at generator level, only the two following Monte Carlo samples, one for the signal and one for the leading background, are considered:

- a sample of ZH events for the signal, where the Higgs boson decays to a bottom quark pair and the Z boson decays to a pair of same flavour opposite sign leptons (electrons, muons or tau leptons)
- a sample of Drell-Yan events, that is the most important background, where a Z boson is produced in association with jets (Z+jets), and decays to a pair of same flavour opposite sign leptons (with invariant mass $m_{ll} > 50$ GeV).

Samples of the same two processes, from pp collisions at $\sqrt{s} = 8$ TeV have been used as well for comparison. Signal and Drell-Yan background events are simulated using the Powheg [80] and MadGraph [81] generators, respectively. The parton showering and hadronization is simulated with Herwig++ [82] for the signal samples and with Pythia 6 [83] and Pythia 8 [84] for the Drell-Yan samples at 8 and 13 TeV, respectively. The samples used and the number of events in each of them are listed in Table 5.1.

Figure 5.10 shows the normalized p_T distributions of the p_T-leading (left) and subleading (right) leptons, for signal and background in the 13 TeV samples, in events with at least two central ($|\eta| < 2.4$) jets with p_T >30 GeV and two central leptons with $p_T > 20$ GeV

Table 5.1 Signal and background samples used for generator level studies at 8 and 13 TeV, with the corresponding number of simulated events.

Energy	Sample	Number of events
8 TeV	/ZH_ZToLL_HToBB_M-125_powheg-herwigpp/	550643
	/DYJetsToLL_M-50_TuneZ2Star_madgraph-pythia6/	1741261
13 TeV	/ZH_HToBB_ZToLL_M-125_powheg-herwigpp/	198566
	/DYJetsToLL_M-50_madgraph-pythia8/	995811

from the Z boson decay. Both for the leading and for the subleading lepton, the p_T spectrum for the signal is harder than the one for Drell-Yan events.

Figure 5.10 Transverse momentum distributions of the p_T leading (left) and subleading (right) lepton, for signal (red) and Drell-Yan (black) simulated events at $\sqrt{s} = 13$ TeV, at generator level. Only events with at least two central jets ($|\eta| < 2.4$) with $p_T > 30$ GeV and two central leptons with $p_T > 20$ GeV from the Z boson decay are considered. Distributions are normalized to unity. Uncertainty is statistical only.

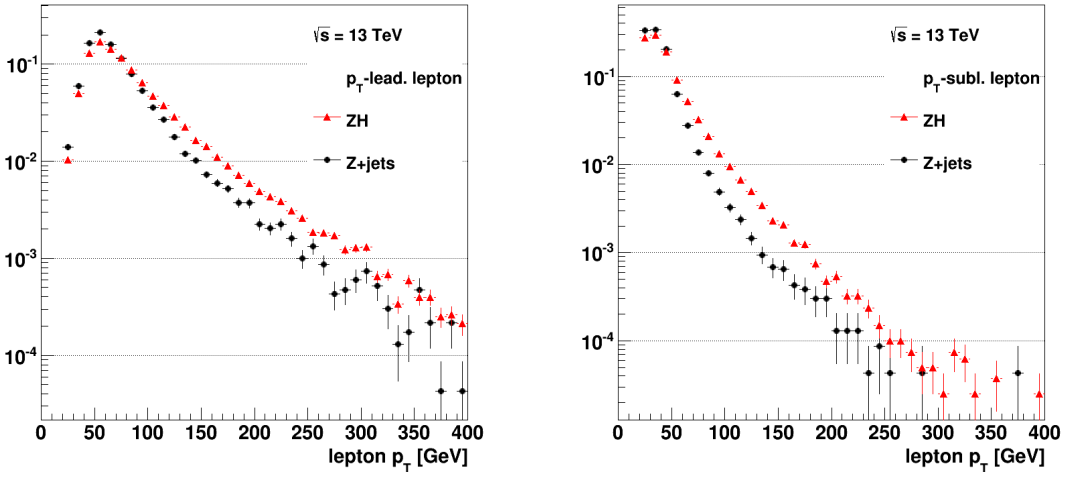
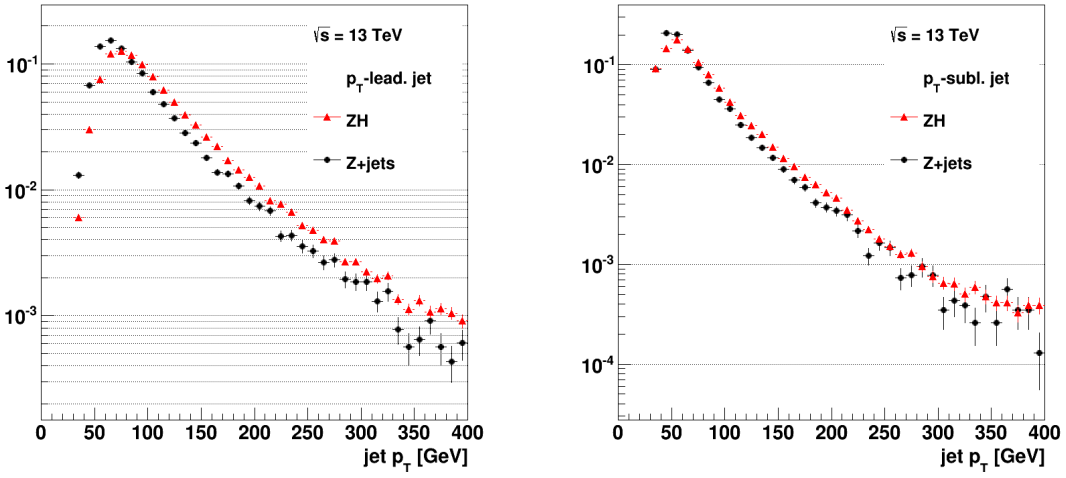


Figure 5.11 shows the normalized p_T distributions of the p_T-leading (left) and subleading (right) jets, for signal and background in the 13 TeV samples, in events with at least two central ($|\eta| < 2.4$) jets with $p_T > 30$ GeV and two central leptons with $p_T > 20$ GeV from the Z boson decay. Similarly to what observed for the leptons, the jet p_T distribution is slightly harder for the signal than for the Drell-Yan background.

Figure 5.12 shows a comparison of kinematic distributions between 8 and 13 TeV generated events. It can be observed that the p_T distribution of leading and subleading jets becomes harder both for the signal and the background when increasing the center of mass energy. In addition to that, the leading and subleading jets in Drell-Yan events are found to be more central at 13 TeV as can be seen from the corresponding η distributions in Figure 5.13. Such difference is instead not observed in signal events.

Figure 5.11 Transverse momentum distributions of the p_T leading (left) and subleading (right) jet, for signal (red) and Drell-Yan (black) simulated events at $\sqrt{s} =$ 13 TeV, at generator level. Only events with at least two central jets ($|\eta| < 2.4$) with $p_T > 30$ GeV and two central leptons with $p_T > 20$ GeV from the Z boson decay are considered. Distributions are normalized to unity. Uncertainty is statistical only.



The cross section for signal and Z+jets background, as a function of the generated p_T of the Z boson, is shown in Figure 5.14. Samples of 8 TeV and 13 TeV events are superimposed for comparison and are normalized to their total production cross sections times branching ratio, according to the values in Table 5.2. A basic selection requiring two central leptons with $p_T > 20$ GeV and two central jets with $p_T > 30$ GeV is applied. The number of events passing this selection are reported in Table 5.3 separately for each sample. The bottom inset of Figure 5.14 shows the signal over background ratios (S/B) as a function of the Z transverse momentum, for the 8 TeV and for the 13 TeV samples. It has the same trend in both cases and increases at high Z p_T, indicating that at 13 TeV the analysis will still be more sensitive in the boosted region, similarly to what was observed from the 8 TeV samples.

Table 5.2 Cross sections and branching ratios for the ZH signal and Drell-Yan background, used for normalization of the plot in Figure 5.14.

Cross sections		
$\sqrt{s}$	ZH $m_H = 125$ GeV	DY Z($\mu\mu$), $m_{ll} > 50$ GeV
8 TeV	0.4153 pb	1177.3 pb
13 TeV	0.8594 pb	2008.4 pb

Branching ratios		
$H \to b\bar{b}$	57.7%	$m_H = 125$ GeV
$Z \to l^+l^-$	10.1%	$l = e, \mu, \tau$

Figure 5.12 Transverse momentum distributions of the p_T leading and subleading jets in Drell-Yan (top) and ZH signal (bottom) simulated events at $\sqrt{s} = 13$ TeV and $\sqrt{s} = 8$ TeV, at generator level. Only events with at least two central jets ($|\eta| < 2.4$) with $p_T > 30$ GeV and two central leptons with $p_T > 20$ GeV from the Z boson decay are considered. Distributions are normalized to unity. Uncertainty is statistical only.

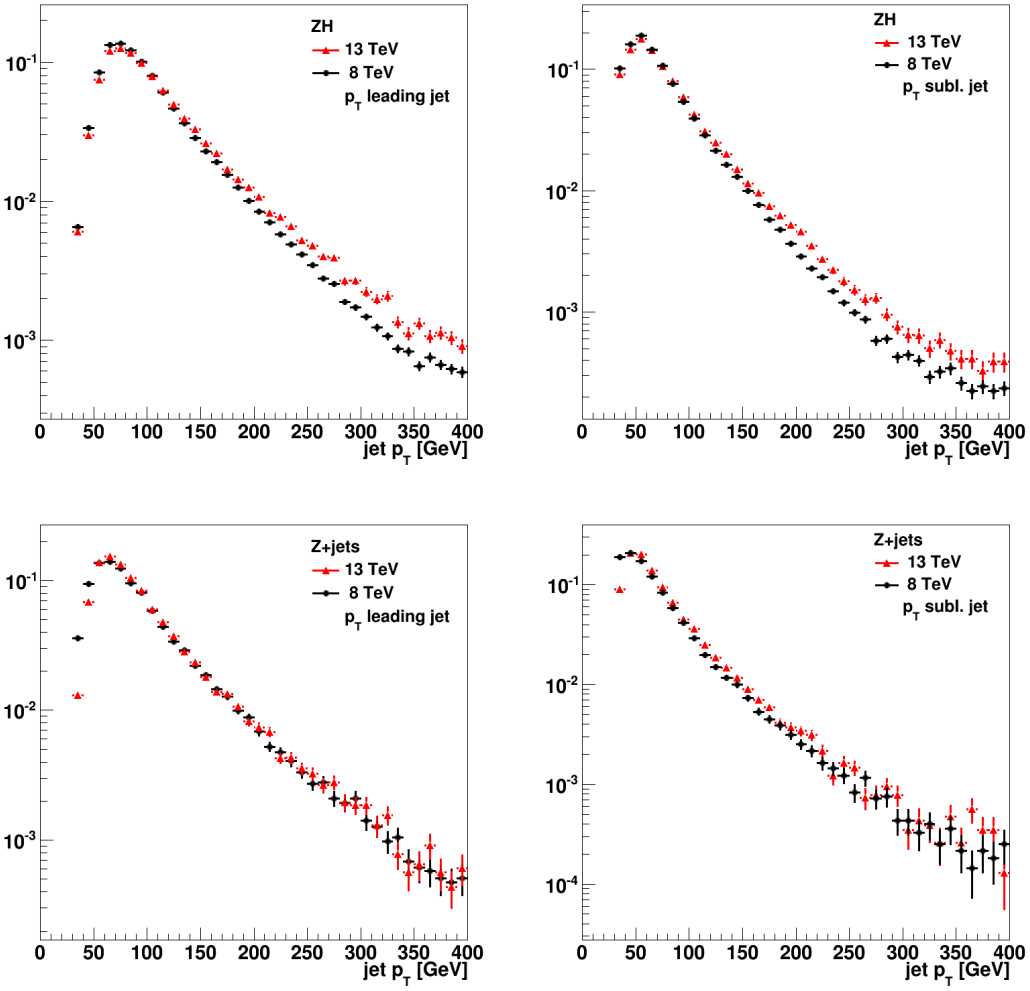
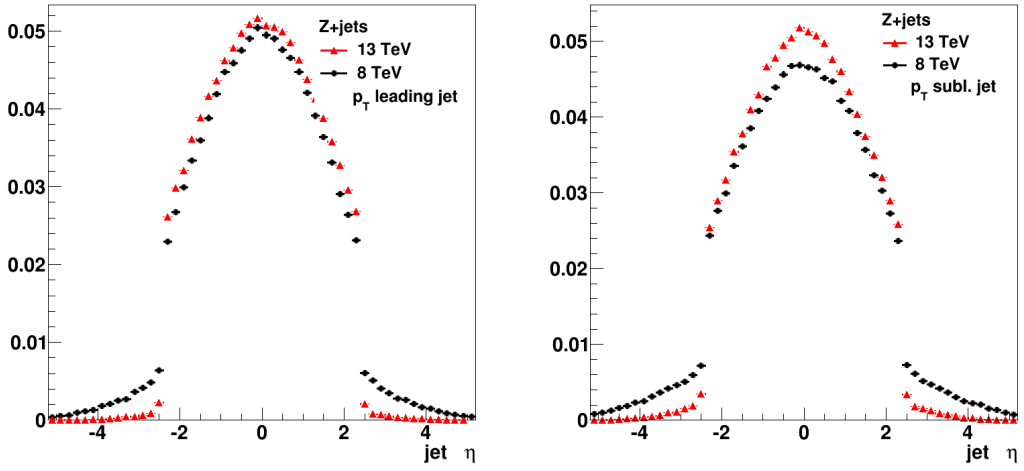


Table 5.3 Number of events passing the basic selection requiring two central leptons with $p_T > 20$ GeV and two central jets with $p_T > 30$ GeV, for signal (ZH) and Drell-Yan background (DY) samples, at 8 and 13 TeV, at generator level.

Energy	Sample	Selected events	Selection efficiency
8 TeV	ZH	243320	44.2%
	DY	27723	1.6%
13 TeV	ZH	80210	40.4%
	DY	23196	2.3%

Figure 5.13 Pseudorapidity distributions of the p_T leading (left) and subleading (right) jets in Drell-Yan simulated events at $\sqrt{s} = 13$ TeV and $\sqrt{s} = 8$ TeV, at generator level. Only events with at least two central jets ($|\eta| < 2.4$) with $p_T > 30$ GeV and two central leptons with $p_T > 20$ GeV from the Z boson decay are considered. Distributions are normalized to unity. Uncertainty is statistical only.

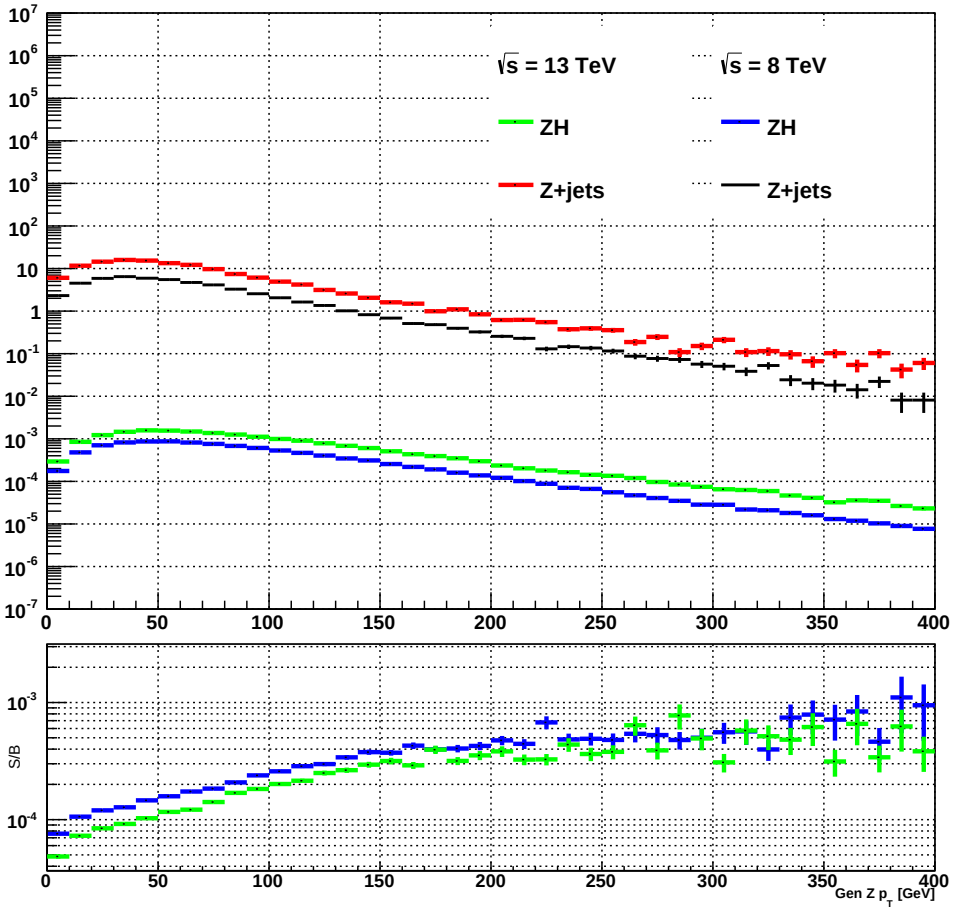


In these preliminary studies, it has been observed that due to the increased center of mass energy, kinematic distributions of the p_T leading and subleading jets become harder and more central. Such differences are more effective for the background than for the signal. Therefore, applying acceptance cuts (requiring at least two central jets and leptons with p_T greater than a given threshold) results in almost the same efficiency for the signal at 13 TeV with respect to the 8 TeV, while the yield of background events passing the selection increases of ∼40%, as can be observed in Table 5.3. For this reason, the larger gain in signal cross section, that is $\sim$ 2 for the signal and $\sim$ 1.7 for the Drell-Yan background with respect to the 8 TeV scenario, does not directly translate in a higher signal over background ratio, as can be seen in the bottom inset of Figure 5.14. The latter plot also shows that the signal over background ratio increases at high p_T of the Z boson, confirming that focusing the analysis in the boosted regime is still convenient at 13 TeV.

For Z p_T ∼100 GeV, the signal over background S/B ratio is found to be lower of about 20% at 13 TeV with respect to the 8 TeV scenario. Based on this, as combined effect of the increased signal cross section but reduced S/B ratio, an increase of $\sim$ 20% on the $S/\sqrt{B}$ ratio is expected at 13 TeV with respect to 8 TeV, for the same integrated luminosity.

In the next Section 5.2.2 results of a simple invariant mass analysis with full simulation Monte Carlo samples at 13 TeV are presented.

Figure 5.14 Cross section [pb] as a function of the generated Z p_T [GeV], for ZH signal and Drell-Yan background events, in the 8 and 13 TeV samples. The bottom inset shows the signal over background ratio in the 8 TeV (blue) and in the 13 TeV (green) samples. Uncertainty is statistical only.



5.2.2 Invariant Mass Analysis on Fully Simulated Samples

A more realistic estimate of the expected Run 2 performances for the Z(ll)H(bb) analysis can be obtained using the available samples that have been processed through a full CMS detector simulation, modeled with GEANT4 [85].

The Run 1 analysis has shown that a multivariate approach based on boosted decision trees gives better results with respect to a cut based invariant mass analysis. Therefore, this approach will likely be adopted also in future analyses. However, a full multivariate analysis is beyond the scope of this Thesis. Instead, a simple cut-based invariant mass analysis has been implemented and the $S/\sqrt{B}$ ratio combined over multiple bins of the

phase space is evaluated. The following paragraphs present the implementation and results of such analysis.

Background Samples Several background processes have been considered:

- production of a Z boson in association with jets from Drell-Yan processes
- production of $t\bar{t}$ pairs
- QCD events, inclusive and enriched in leptons
- production of jets in association with photons

They refer to a scenario of pp collisions with 25 ns bunch crossing and an average pile up of 20 interactions. The full list of background samples is reported in Table 5.4, together with the cross sections used to normalize the distributions in the following plots. The total number of generated events before any selection and entering the normalization is also reported.

Signal Reconstruction The reconstruction of the Z(ll)H(bb) signal implies the reconstruction of the Z and the Higgs boson. A pair of same flavour, opposite sign leptons in the tracker acceptance $|\eta| < 2.4$ are needed to reconstruct a Z candidate. The Higgs boson candidate is obtained from the two p_T leading jets in the event, requiring that they are central and with a transverse momentum of at least 15 GeV. Events are classified in five categories depending on the vector boson decay mode: Z($\mu\mu$), Z(ee), W($\mu\nu$), W($e\nu$) and Z($\nu\nu$). Generated signal events include only Z(ll) decays, with $l = e, \mu, \tau$. However, it can happen that one of the leptons is not detected because it lies out of the detector acceptance. In this case, Z(ll) events can be reconstructed as W($l\nu$) events, or even as Z($\nu\nu$) events when both leptons are not detected. The number of events reconstructed in each final state, as a function of generated signal events are provided in Table 5.5. For the Z($\mu\mu$) channel, the fraction of correct assigned events is approximately 73%, while $\simeq$20% of the events are misreconstructed as W($\mu\nu$) events. For the Z(ee) final state, the probability of correct assignment is 60%, with $\simeq$35% of generated events that are reconstructed in the W($e\nu$) channel.

Signal Region Definition To enhance the sensitivity, the signal region is divided into three categories, according to the transverse momentum of the Z boson candidate. The following boost regions are defined:

- a low-p_T bin, including events with Z p_T between 50 and 100 GeV
- a medium-p_T bin, for events with Z p_T between 100 and 150 GeV
- a high-p_T bin, in which Z p_T >150 GeV

Table 5.4 List of all background samples used, together with their cross section and total number of events used for normalization.

Background samples

	σ [pb]	Events
Drell Yan		
DYJetsToLL_M-50_13TeV-madgraph-pythia8	6025.2	2829156
DYJetsToLL_M-50_HT-100to200_Tune4C_13TeV-madgraph-tauola	194.3	4054144
DYJetsToLL_M-50_HT-200to400_Tune4C_13TeV-madgraph-tauola	52.24	4666479
DYJetsToLL_M-50_HT-400to600_Tune4C_13TeV-madgraph-tauola	6.546	4931352
DYJetsToLL_M-50_HT-600toInf_Tune4C_13TeV-madgraph-tauola	2.179	4493550
$t\bar{t}$		
TT_Tune4C_13TeV-pythia8-tauola	809.1	25446877
γ+jets		
GJets_HT-100to200_Tune4C_13TeV-madgraph-tauola	1534	4734221
GJets_HT-200to400_une4C_13TeV-madgraph-tauola	489.9	4533397
GJets_HT-400to600_Tune4C_13TeV-madgraph-tauola	62.05	4560788
GJets_HT-600toInf_Tune4C_13TeV-madgraph-tauola	20.87	4341164
QCD		
QCD_HT-100To250_13TeV-madgraph	28730000	4123591
QCD_HT-250To500_13TeV-madgraph	670500	2668164
QCD_HT-500To1000_13TeV-madgraph	26740	4063331
QCD_HT-1000ToInf_13TeV-madgraph	769.7	1464447
QCD, e/γ enriched		
QCD_Pt-10to20_EMEnriched_Tune4C_13TeV_pythia8	1263834000	2000075
QCD_Pt-20to30_EMEnriched_Tune4C_13TeV_pythia8	4741100	1987120
QCD_Pt-30to80_EMEnriched_Tune4C_13TeV_pythia8	10410400	5138543
QCD_Pt-80to170_EMEnriched_Tune4C_13TeV_pythia8	557582	1959500
QCD, μ enriched		
QCD_Pt-30to50_MuEnriched_Tune4C_13TeV_pythia8	1340688	3745697
QCD_Pt-50to80_MuEnriched_Tune4C_13TeV_pythia8	331948.2	3896706
QCD_Pt-80to120_MuEnriched_Tune4C_13TeV_pythia8	72695.8	3835952

Table 5.5 Number of signal events generated and reconstructed in each final state.

Decay mode	Reconstructed				
Generated	Z($\mu\mu$)	Z(ee)	W($\mu\nu$)	W($e\nu$)	Z($\nu\nu$)
Z($\mu\mu$)	26042	-	8318	23	1466
Z(ee)	-	19572	6	11930	1452
Z($\tau\tau$)	320	198	3459	2674	14885

Events with Z p_T smaller that 50 GeV are not considered due to the very low signal over background ratio. Some cuts on the p_T and on the b-tagging discriminator of the two jets used to reconstruct the Higgs boson have been applied, as well as on the invariant mass of the dilepton system to be compatible with a Z decay. Due to the higher boost, the jets from the Higgs decay are expected to be close to each other for events in the high-p_T bin, therefore a cut on the angular distance ΔR between the two jets is applied. Moreover a cut on the missing energy is applied, as no invisible particles are expected in signal events in the dielectron and dimuon channels. Two different thresholds on the b-tagging discriminator are applied to the most and least b-tagged jets used to reconstruct the Higgs candidate. The chosen b-tagging discriminator is the output of the CMVA algorithm, a tagger that takes into account information about tracks, secondary vertices and soft leptons inside the jet, combining them in a multivariate analysis. With respect to the CSV algorithm used for Run analyses, the CMVA algorithm provides a better c-jet rejection. Similarly to the CSV tagger, the CMVA output lies in the range [0,1] and the distribution for b-jets is shifted towards higher values with respect to other jets.

All cuts applied to select the signal region are summarized in Table 5.6, while the number of signal events passing after each step of this selection and the relative efficiencies are reported in Tables 5.7 and 5.8 for the muon and electron channels respectively.

Plots in Figure 5.15 show the invariant mass of the two p_T leading jets, i.e. of the Higgs boson candidate, after applying the signal region selection, separately for the Z(ee) and Z($\mu\mu$) channels and for the three p_T bins. The signal and background yields refer to an integrated luminosity of 19 fb^{-1}. Drell-Yan events have been classified in three disjoint categories, according to the flavour of the partons originating the jets used in the Higgs boson reconstruction:

- if none of the jets comes from a b-quark, the event enter the **Z+udcsg** category
- if only one of the two jets is originated by a b-quark, the event is labeled as **Z+b**
- events where both jets are from a b-quark enter the **Z+bb** category

The Z+bb component turns out to be the most important background in the signal region because its topology is very similar to the signal.

The figure of merit S/$\sqrt{\text{B}}$ is used here and in the following of this Thesis to estimate the sensitivity of the analysis in a given region, S and B being the number of signal

Table 5.6 Definition of the signal region. All cuts applied to reconstructed events are listed below.

Signal region selection

Variable	Cut	
$p_T(j_1, j_2)$	$(> 20, > 20)$ GeV	j_1, j_2: jets used to reconstruct the Higgs candidate
Z mass	75 - 105 GeV	
CMVA (j_1, j_2)	$(> 0.76, > 0.7)$	b-tagging discriminator from the CMVA algorithm
MET	< 60 GeV	
$\Delta R(j_1, j_2)$	< 1.6	only in the high boost region (Z $p_T > 150$ GeV)

and background events. Such figure of merit is evaluated summing in quadrature the contributions of the three signal enriched mass bins (m(jj) = [70-100], [100-130] and [130-160] GeV) for each of the three p_T regions, resulting in a combination of 9 bins for each of the two decay channels. The signal and total background event yields in the three signal enriched mass bins m(jj) = [70-100], [100-130] and [130-160] GeV are provided in Table 5.9, separately for the two considered channels. Among the three boost regions, the high-p_T region is found to be the most sensitive. After applying the signal region selection, the total $S/\sqrt{B}$ ratio for the muon (electron) channel, combined over the three mass bins and three p_T bins, is $\sim$0.65 ($\sim$0.59) standard deviations for a luminosity of 19 fb^{-1}. Systematic uncertainties are not taken into account at this stage, because object reconstruction efficiencies, energy scales and the newest theory predictions have not yet been evaluated in the Run 2 scenario, and all measurements are supposed to be uncorrelated. Moreover, the contribution of diboson production background has not been considered as full simulation samples for early Run 2 studies are not available at the time of writing this Thesis. Basing on the Run 1 analysis, the contribution of the Z(ll)Z(bb) channel is expected to be of the order of 2-20% in the mass bin from 70 to 100 GeV.

Background Control Regions The definition of control regions enriched in background events is necessary for data driven estimate of the most important background processes. For this analysis, three different control regions are defined:

- **Z+Light Flavour** jets for Drell-Yan events with no b-jets
- **Z+Heavy Flavour** jets for Drell-Yan events with one or more b-jets
- $\mathbf{t\bar{t}}$ for top quark pair production

where the flavour categorization is applied on the two jets used for the Higgs boson reconstruction and relies on Monte Carlo truth information. A good control region

Table 5.7 Number of signal events passing the signal region selection as in Table 5.6 and relative efficiency with respect to the previous cut, for the Z($\mu\mu$) channel. The total efficiency after all selections, with respect to the total number of signal events, is also provided. Numbers are given separately for the three p_T-bins.

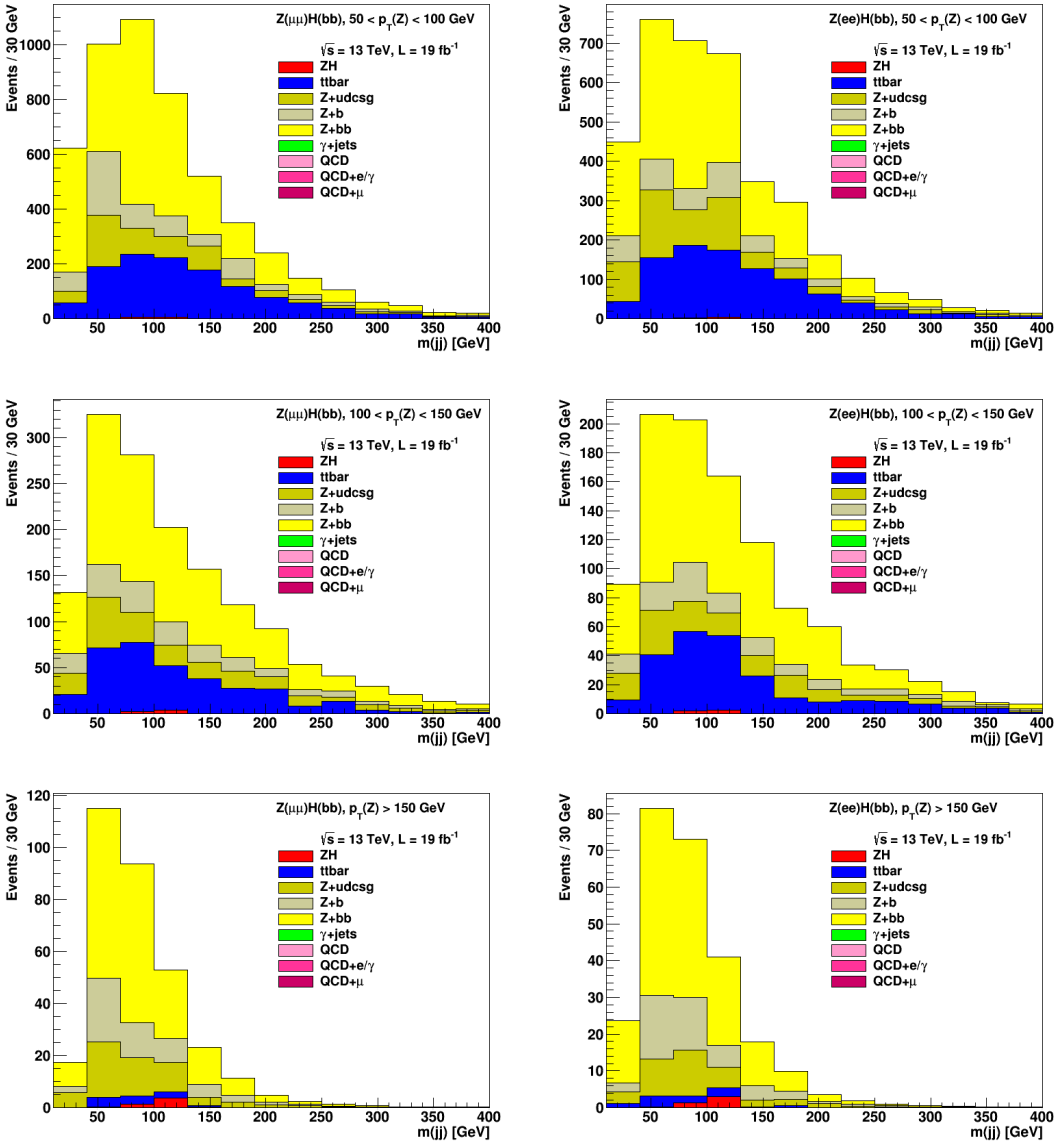
Muon channel			
Selection	Total events	Selected events	Efficiency [%]
Z$\rightarrow \mu\mu$	198566	26362	13.3
Z mass = $[75-105]$ GeV	26362	24971	94.7
Low p_T bin			
p_T(Z) = $[50-100]$ GeV	24971	8476	33.9
$p_T(j_1, j_2) > (20, 20)$ GeV	8476	8476	100
CMVA (j_1) >0.76	8476	5932	70.0
CMVA (j_2) >0.7	5932	2399	40.4
MET < 60 GeV	2399	2284	95.2
Total efficiency			1.15
Intermediate p_T bin			
p_T(Z) = $[100-150]$ GeV	24971	5261	21.1
$p_T(j_1, j_2) > (20, 20)$ GeV	5261	5261	100
CMVA (j_1) >0.76	5261	3697	70.3
CMVA (j_2) >0.7	3697	1489	40.3
MET < 60 GeV	1489	1373	92.2
Total efficiency			0.69
High p_T bin			
p_T(Z) > 150 GeV	24971	5220	20.9
$p_T(j_1, j_2) > (20, 20)$ GeV	5220	5220	100
CMVA (j_1) >0.76	5220	3705	71.0
CMVA (j_2) >0.7	3705	1646	44.4
MET < 60 GeV	1646	1322	80.3
$\Delta R(j_1, j_2)$	1322	1190	90.0
Total efficiency			0.60

Table 5.8 Number of signal events passing the signal region selection as in Table 5.6 and relative efficiency with respect to the previous cut, for the Z(ee) channel. The total efficiency after all selections, with respect to the total number of signal events, is also provided. Numbers are given separately for the three p_T-bins.

Electron channel

Selection	Total events	Selected events	Efficiency [%]
Z$\rightarrow ee$	198566	19770	10.0
Z mass $= [75 - 105]$ GeV	19770	18851	95.3
Low p_T bin			
p_T(Z) $= [50 - 100]$ GeV	18851	6173	32.7
$p_T(j_1, j_2) > (20, 20)$ GeV	6173	6173	100
CMVA (j_1) >0.76	6173	4297	69.6
CMVA (j_2) >0.7	4297	1686	39.2
MET $<$ 60 GeV	1686	1616	95.8
Total efficiency			0.86
Intermediate p_T bin			
p_T(Z) $= [100 - 150]$ GeV	18851	3955	21.0
$p_T(j_1, j_2) > (20, 20)$ GeV	3955	3955	100
CMVA (j_1) >0.76	3955	2794	70.6
CMVA (j_2) >0.7	2794	1117	40.0
MET $<$ 60 GeV	1117	1004	89.9
Total efficiency			0.51
High p_T bin			
p_T(Z) $>$ 150 GeV	18851	4233	22.5
$p_T(j_1, j_2) > (20, 20)$ GeV	4233	4233	100
CMVA (j_1) >0.76	4233	2976	70.3
CMVA (j_2) >0.7	2976	1290	43.3
MET $<$ 60 GeV	1290	1064	82.5
$\Delta R(j_1, j_2)$	1064	972	91.4
Total efficiency			0.49

Figure 5.15 Reconstructed invariant mass distributions of the two p_T leading jets, i.e. of the Higgs boson candidate after applying the signal region selection according to Table 5.6. Left: Z($\mu\mu$) channel, right: Z(ee) channel, separately for each of the three boost regions, in order of increasing Z p_T from top to bottom.



definition aims at high purity in the addressed process. The selection cuts used for the 8 TeV analysis still provide a good separation of the background processes on the 13 TeV samples, with the exception of the Z+Heavy Flavour region, that suffers from contamination of $t\bar{t}$ events. It is due to the larger $t\bar{t}$ production cross section, that increases of a factor 4 from 8 to 13 TeV, while the same scale factor for Drell-Yan production is only ∼2. The higher jet multiplicity in $t\bar{t}$ events provides an handle to reject such events in the Z+Heavy Flavour control region. Therefore, events with additional jets with $p_T > 25$ GeV are vetoed. The full list of cuts applied to define the three control regions is reported in Table 5.10. The definition of control regions is the same regardless of the electron or muon final state and the p_T bin.

Table 5.9 Signal (sig) and background (bkg) yields after applying the signal region selection provided in Table 5.6, in the three signal enriched mass bins, in each of the three p_T-bins, for the muon and electron channels separately. The $S/\sqrt{B}$ ratio is also reported, separately for each mass bin and combined in each boost region.

Muon channel

m(jj)	50 < Z p_T < 100 GeV			100 < Z p_T < 150 GeV			Z p_T > 150 GeV		
[GeV]	Sig	Bkg	$S/\sqrt{B}$	Sig	Bkg	$S/\sqrt{B}$	Sig	Bkg	$S/\sqrt{B}$
70-100	4.1	1100	0.123	2.1	280	0.125	1.4	90	0.149
100-130	5.5	820	0.192	3.6	200	0.254	3.6	50	0.514
130-160	0.5	520	0.023	0.3	160	0.025	0.3	25	0.067
Combined	**0.229 ± 0.005**			**0.285 ± 0.008**			**0.540 ± 0.013**		

Electron channel

m(jj)	50 < Z p_T < 100 GeV			100 < Z p_T < 150 GeV			Z p_T > 150 GeV		
[GeV]	Sig	Bkg	$S/\sqrt{B}$	Sig	Bkg	$S/\sqrt{B}$	Sig	Bkg	$S/\sqrt{B}$
70-100	2.7	700	0.102	1.7	200	0.121	1.3	70	0.152
100-130	4.0	670	0.156	2.5	160	0.197	3.0	40	0.481
130-160	0.4	350	0.020	0.2	120	0.016	0.2	20	0.052
Combined	**0.187 ± 0.005**			**0.233 ± 0.007**			**0.507 ± 0.013**		

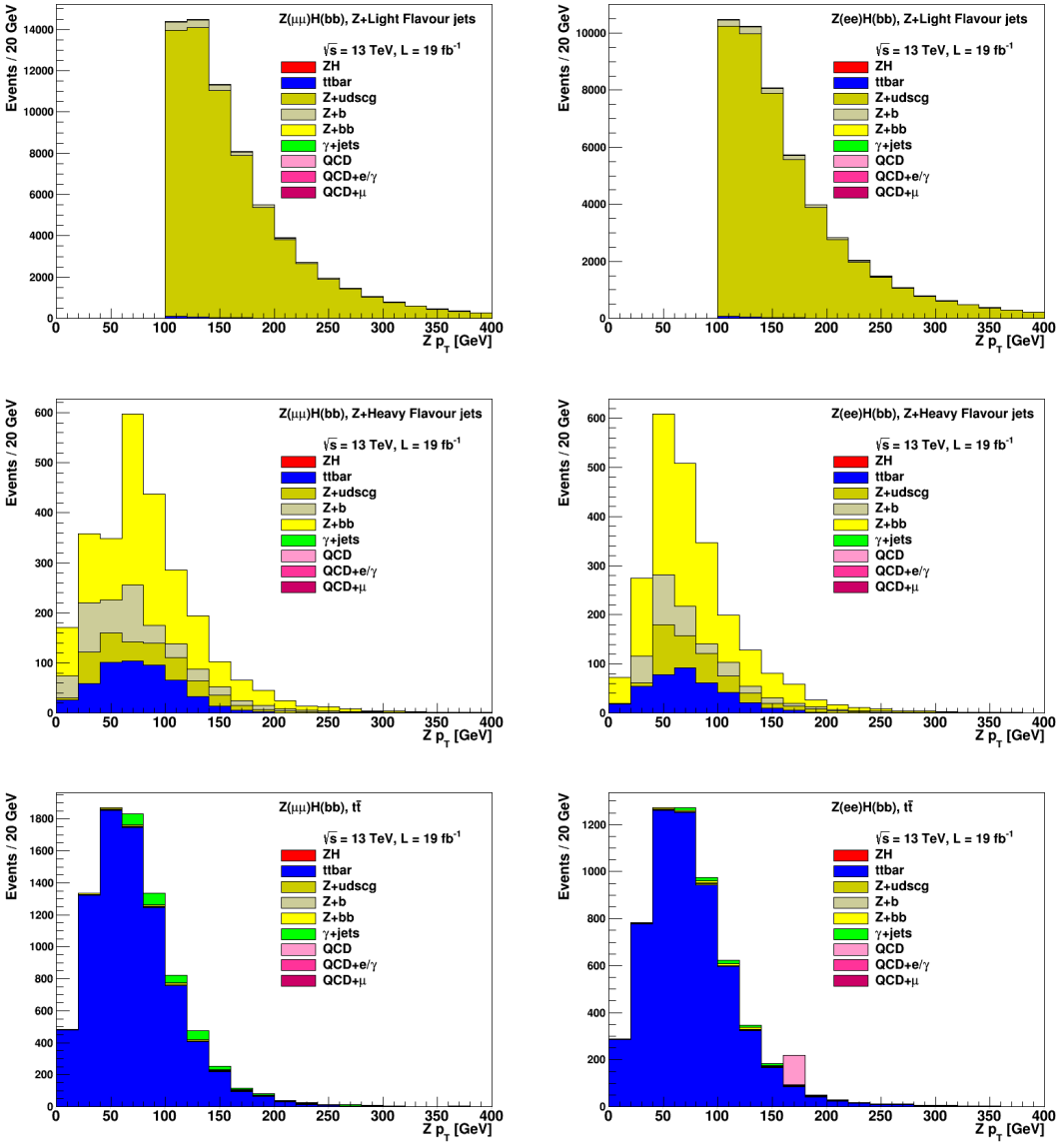
Figure 5.16 shows the Z p_T distributions in the three control regions, separately for the electron and muon channels.

Table 5.10 Definition of the control regions for Drell-Yan Z+Light Flavour, Drell-Yan Z+Heavy Flavour and $t\bar{t}$ backgrounds. All cuts applied to reconstructed events are listed below.

	Control Region		
Variable	**Z+LF**	**Z+HF**	**$t\bar{t}$**
Higgs p_T	> 100 GeV	—	> 100 GeV
Higgs mass	—	veto 90-150 GeV	—
Z p_T	> 100 GeV	—	—
Z mass	70-110 GeV	70-110 GeV	veto 70-110 GeV
CMVA (j_1, j_2)	$(< 0.76, < 0.76)$	$(> 0.76, > 0.7)$	$(> 0.76, > 0.7)$
N add. jets	< 2	< 1	—
$\lvert\Delta\phi$(Z,H)$\rvert$	> 2.9	> 2.9	—

The sensitivity of the analysis has been evaluated through the $S/\sqrt{B}$ figure of merit. Neglecting correlations and systematic uncertainties, the significance combined over the three p_T regions is found to be ∼0.65 σ for the muon channel and ∼0.59 σ for the electron channel, resulting in a overall significance of ∼0.88 σ for a luminosity of 19 fb^{-1}. Diboson

Figure 5.16 Distributions of the Z transverse momentum in the Z+Light Flavour jets (top), Z+Heavy Flavour jets (middle) and $t\bar{t}$ (bottom) control regions, applying the cuts in Table 5.10. Plots for the muon (electron) channel are on the left (right) coloumn.



processes have not been included in the background due to the lack of fully simulated samples for Run 2 studies. However, it can be roughly estimated that of the contribution of the Z(ll)Z(bb) process to the background would decrease the $S/\sqrt{B}$ of approximately 5%, from 0.88 to 0.85 σ. An evaluation of systematic uncertainties for Run 2 is not yet available. When taking into account the same systematic uncertainty as in Run 1 analysis, the $S/\sqrt{B}$ ratio would further decrease from 0.85 to 0.79 σ . When comparing the latter result with the $\sim$ 0.60 σ obtained in the 8 TeV invariant mass analysis for the same integrated luminosity, it translates in a 30% improvement in sensitivity for the 13 TeV analysis. Such improvement matches the expectations from the preliminary kinematic study presented in Section 5.2.1.

5.2.2.1 B-tagging Optimization

In the results presented so far, the discriminator from the CMVA algorithm has been used for b-tagging, applying the selection cuts listed in Table 5.6 to identify the signal region. However, many other b-taggers are available and can in principle be used. The ability of identifying b-jets is of primary importance for the VH,H(bb) analysis, thus it is worth to actually compare the performances of the different algorithms and identify the most suitable. The following b-taggers have been considered:

- Jet Probability (JP)
- Jet B Probability (JBP)
- CMVA
- Combined Secondary Vertex (CSV)

The idea behind the JP, JBP and CSV algorithms has been described in Chapter 4, while the CMVA algorithm has been introduced in the previous Section 5.2.2. Concerning the CSV algorithm, the last improved version [56] developed for Run 2 has been used for the following studies. It relies on a different approach to reconstruct secondary vertices: they are reconstructed independently of jets, starting from a collection of tracks with a large impact parameter with respect to the primary vertex. The second main difference is the use of a neural network to combine information from the different variables. The overall effect is an enhanced b-tagging efficiency in events with high-p_T jets and in large pile up environments.

In order to compare the performances of the above mentioned taggers, samples of simulated events at 13 TeV has been used, for the Z(ll)H(bb) signal and for the Drell-Yan and $t\bar{t}$ backgrounds, that turn out to be the most relevant ones.

The comparison is based on the figure of merit $S/\sqrt{B}$, S and B being the total number of signal and background events in a given bin of the reconstructed Higgs mass distribution. The $S/\sqrt{B}$ ratio is combined summing in quadrature the values of the three signal enriched mass bins, namely m(jj) = $[70, 100], [100, 130]$ and $[130, 160]$ GeV. For each channel, a single $S/\sqrt{B}$ ratio is obtained combining the values obtained in the three p_T bins. To estimate the global performance without distinguishing between the muon and the electron channels, the values for the single channels are summed in quadrature.

Search for Optimal Cuts For each b-tagger, the $S/\sqrt{B}$ figure of merit is calculated applying the signal region selection listed in Table 5.6, except for the b-tagging selection on the two Higgs jets. Cuts on b-tagging discriminator are varied independently for the most and least b-tagged jets used to reconstruct the Higgs candidate. The figure of merit is calculated for every combination of cuts, resulting in a two dimensional scan. A first coarser scan to identify an optimal region is followed by a finer scan, from which the

best cuts can be read. The cuts corresponding to the maximum $S/\sqrt{B}$ value represent the optimal choice. In general, optimal cuts are found to be slightly different in the two channels, although within the statistical uncertainty of the samples. The plots in Figure 5.17, that shows the fine 2D scan for the CMVA tagger, separately for the Z($\mu\mu$) and Z(ee) channels. If it is convenient to have the same signal region selection for both Z($\mu\mu$) and Z(ee) channels, cuts should be chosen that give the highest combined $S/\sqrt{B}$ ratio. Plots in Figure 5.18 show the combined Z(ee)+Z($\mu\mu$) scans for the CSV and CMVA taggers. Operating points for different algorithms are listed in Table 5.11.

The results for all different algorithms taken into account are provided in Table 5.12, separately and combined over the two channels. For additional comparison, the version of the CSV algorithm used for the Run 1 analysis has been also considered. It is labeled as CSVV0 and the results are given both when using the same cuts as in the 8 TeV analysis (CSVV0 Run 1 cut) and after optimization on the 13 TeV samples. Using the new CSV algorithm developed for Run 2 instead of the Run 1 algorithm leads to an increase of approximately 6% on the Z(ll)H(bb) sensitivity. The use of multivariate techniques, exploited by the CSV and CMVA taggers, allows to better identify b-jets with respect to simpler algorithms based on track parameters only, e.g. the Jet Probability algorithms. The most suitable choice for the future analysis turns out to be the new CSV algorithm, applying asymmetric cuts on the two Higgs jets. The best cuts correspond to a b-tagging efficiency of ∼75% (CSV discriminator > 0.68) and ∼50% (CSV discriminator > 0.92) for the two jets. With this optimized choice, a $S/\sqrt{B} \sim 0.92\,\sigma$ is expected to be achieved with 19 fb^{-1} of data when combining the two Z(ll) channels, to be compared with a maximum $S/\sqrt{B} \sim 0.89\,\sigma$ that can be obtained when using the CMVA algorithm. Optimizing the b-tagging choice leads to a sensitivity improvement of approximately 5% with respect to the one reported in Table 5.6.

Figure 5.17 Fine scans on the CMVA discriminator cuts, for the Z($\mu\mu$) (top) and Z(ee) (bottom) channels. The value printed on each bin is the S/$\sqrt{B}$ ratio combined over the three p_T-bins, as explained in Section 5.2.2.1. The optimal selections in the muon and electron channels are (0.72, 0.80) and (0.68, 0.78) respectively.

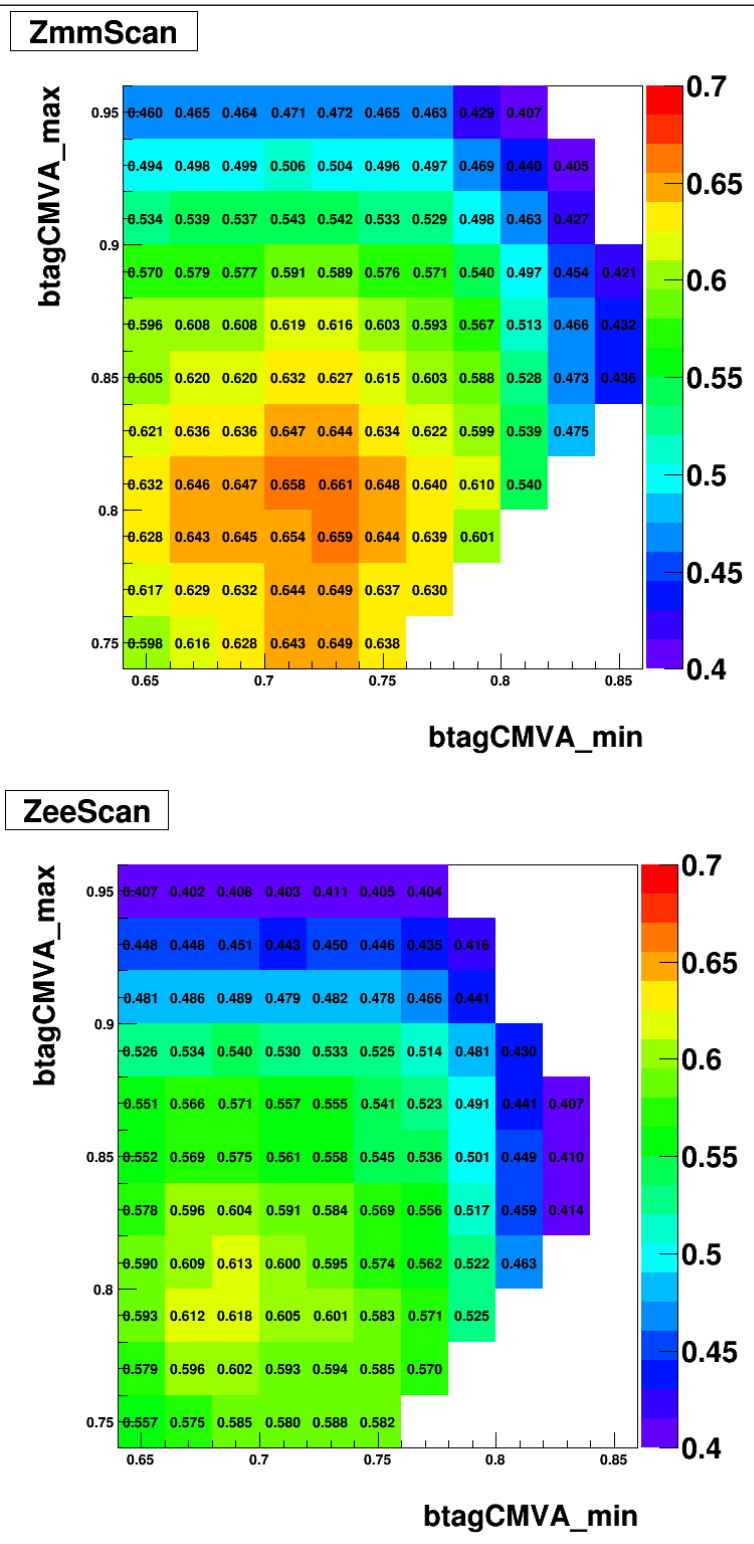


Figure 5.18 Fine scans on the CMVA (top) and CSV (bottom) discriminator cuts, combined for the Z($\mu\mu$) and Z(ee) channels. The best cuts are (0.68, 0.92) for the CSV tagger and (0.68, 0.78) for the CMVA tagger. When applying this selection, the CSV algorithm gives a $S/\sqrt{B} = 0.92$, while the same value for the CMVA algorithm is 0.89.

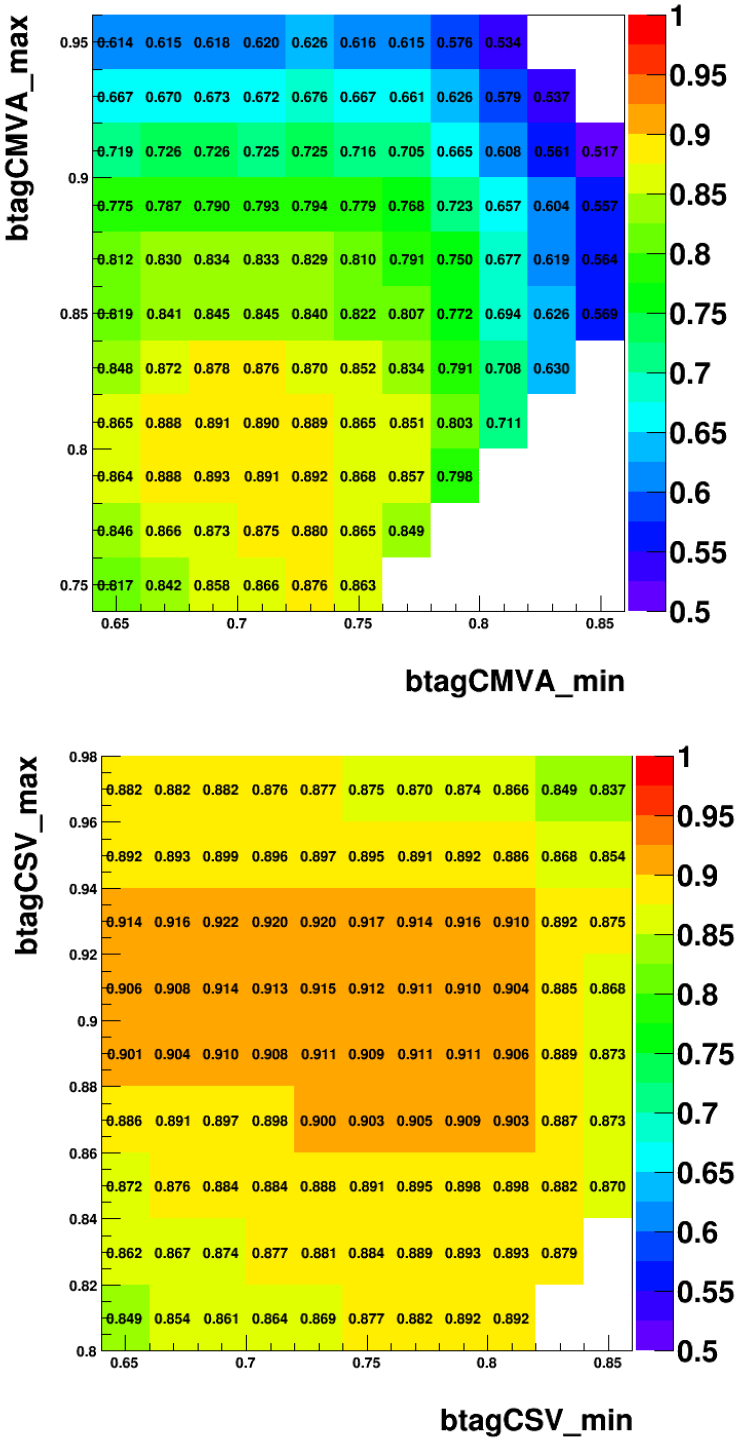


Table 5.11 Working points corresponding to a given inclusive b-jet efficiency for the CSV and CMVA algorithms, measured from a sample of simulated $t\bar{t}$ event at $\sqrt{s} = 13$ TeV.

b-jet efficiency	**Discriminator cut**			
[%]	CSV	CMVA	JP	JBP
50	0.92	0.80	0.68	3.30
55	0.90	0.78	0.61	3.00
60	0.87	0.76	0.57	2.70
65	0.83	0.74	0.50	2.44
70	0.77	0.71	0.43	2.14
75	0.66	0.68	0.37	1.84

Table 5.12 Best cuts on the most (max) and least (min) b-tagged jets assigned to the Higgs candidate, together with the corresponding $S/\sqrt{B}$ ratio. Results are shown for all considered algorithms, separately and combined for the muon and electron channels.

Tagger	**Best cuts**								
	Muon channel			Electron channel			Channels combined		
	min	max	$S/\sqrt{B}$	min	max	$S/\sqrt{B}$	min	max	$S/\sqrt{B}$
CSVV0 (Run 1 cuts)	0.50	0.90	0.645	0.50	0.90	0.591	0.50	0.90	0.86
CSVV0	0.60	0.85	0.654	0.60	0.85	0.590	0.60	0.85	0.88
CSV	0.68	0.92	0.693	0.78	0.92	0.618	0.68	0.92	0.92
CMVA	0.72	0.80	0.661	0.68	0.78	0.618	0.68	0.78	0.89
JP	0.50	0.62	0.622	0.50	0.70	0.572	0.50	0.66	0.84
JBP	2.60	2.80	0.592	2.00	3.40	0.553	2.00	3.40	0.80

Chapter 6

Effect of Detector Upgrades on the Z(ll)H(bb) Sensitivity

The aim of the work presented in this Chapter is to estimate the b-tagging performance in future upgrade scenarios, studying the case of the Z(ll)H(bb) channel. This goal is achieved deriving a parametrization of the b-tagging performance for the different upgrade scenarios considered. Finally, signal and background events are scaled accordingly to obtain the prediction on the analysis performance.

6.1 Detector Upgrade Scenarios

The following five scenarios have been implemented in the complete detector simulation and have been taken into account:

- the present detector (Phase 0), taking data of proton collisions at an energy of 13 TeV, with average pile up of 50 interactions per bunch crossing. On this sample, a closure test of the parametrization is performed, as explained in Section 6.3
- the detector after Phase 1 upgrades, taking data of proton collisions at $\sqrt{s} = 14$ TeV
  - neglecting aging effects and in an ideal case of no pile up
  - neglecting aging effects in a 50 pile up environment
  - considering the aging effect of the ECAL and outer tracker systems due to 1000 fb^{-1} of delivered luminosity, in high pile up environment with an average of 140 interactions per bunch crossing. No aging is considered for the pixel tracker because it is planned to be replaced. This scenario is used to estimate the performance of the detector in the high luminosity phase of the LHC and motivates the need for the Phase 2 upgrade

- the detector after Phase 2 upgrades in the high pile up (PU = 140) environment. An aging equivalent to 1000 fb^{-1} delivered luminosity is accounted for the ECAL in the barrel region, as it is not planned to be replaced. This situation provides a benchmark to estimate the performance of the Phase 2 detector. Concerning the pixel tracker considered for this scenario, it is assumed to be identical to one adopted in the Phase 1 upgrade, because the final Phase 2 layout is not yet decided. However, the pixel size will likely be smaller in the Phase 2 than the Phase 1 pixel tracker. Presently, pixel cells of 25×100 and 50×50 μm^2 are being considered for the Phase 2 pixel detector. A better hit resolution, achieved thanks to a smaller pixel size, will improve the quality of reconstructed tracks and hence improve b-tagging performances.

The sensitivity of the analysis in the different scenarios is again evaluated by the $S/\sqrt{B}$ ratio, in the same way as explained in Chapter 5. In order to highlight the enhancement in sensitivity due to the upgrade of the detector, results presented in the following paragraphs do not take into account the cross section increase of about 10% when passing from 13 to 14 TeV center of mass energy.

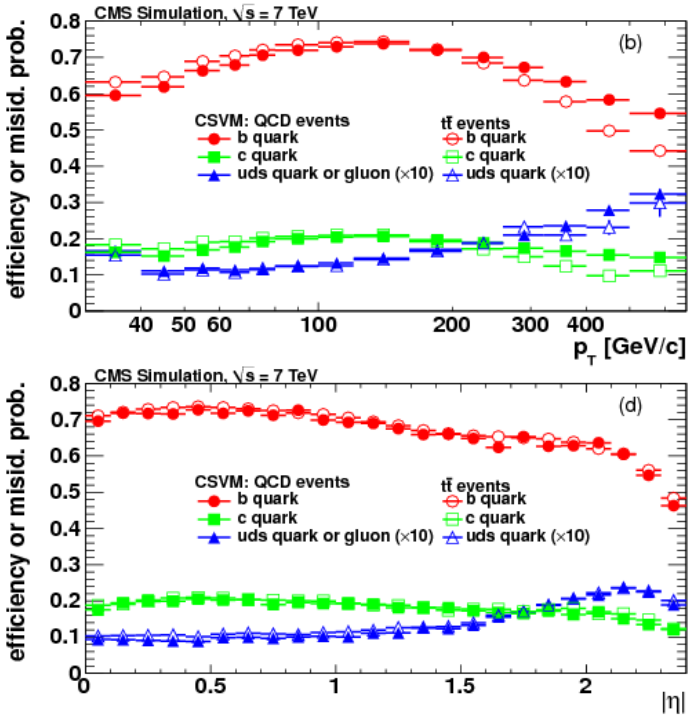
As full simulation samples for the ZH signal and main backgrounds in future upgrade scenarios are not yet available, a parametrized approach has been adopted, using fully simulated $t\bar{t}$ events in the future scenarios mentioned in the previous Section. Such $t\bar{t}$ samples have been used to derive a parametrized b-tagging efficiency (and mistag rates) as a function of the jet p_T and η, which has been propagated to the 13 TeV signal and background samples available. In order to estimate the accuracy of this method, the b-tagging efficiency parametrization has been derived for the 13 TeV samples and a closure test has been performed, comparing results obtained using the actual b-tagging discriminator to the ones obtained applying the parametrized efficiency derived from a $t\bar{t}$ sample.

6.2 B-tagging Efficiency Parametrization

The performance of b-jet identification depends on the momentum and direction of the jet. For a given cut on a b-tagging discriminator, the b-tagging efficiency and misidentification probability can be studied as a function of the p_T and η of the jet. Figure 6.1 shows the b-tagging efficiency (and mistag rates) as a function of the jet p_T and η for the medium working point of the CSV algorithm used in Run 1 analyses [69], corresponding to a nominal inclusive b-efficiency of 60% and a mistag rate for light jets of 1%. For the CSV algorithm, the b-efficiency and c-jet mistag rate increase up to $p_T \sim 100$ GeV and decrease at very high p_T >200 GeV. This happens because at very low transverse momentum, the resolution on the track impact parameter is spoiled by the effect of multiple scattering, while at very high transverse momentum the decrease in efficiency is caused by the increased radius of curvature of tracks and decay length of B hadrons. The mistag rates for light flavour (u,d,s,g) jets increases with the jets transverse

momentum, due to the higher particle multiplicity and number of overlapping hits in the first layers of the tracker. The b-efficiency slightly degrades in the forward η region.

Figure 6.1 B-tagging efficiency and mistag rates as a function of the jet p_T (top) and η (bottom) for the medium working point of the CSV algorithm used in Run 1 analyses, in simulated samples of $t\bar{t}$ and QCD events at $\sqrt{s} = 7$ TeV [69].



The b-tagging efficiency and mistag rates has been studied differentially in bins of (p_T, η), after selecting the discriminator threshold corresponding to a given inclusive b-jet efficiency. An inclusive sample of $t\bar{t}$ events at 13 TeV has been used because it is naturally enriched in b-jets from top decays. The new version of the CSV and the CMVA algorithms have been used, as they show the most promising performances for the Z(ll)H(bb) analysis. Discriminator cuts and corresponding inclusive b-efficiencies for these two taggers are listed in Table 5.11. B-tagging efficiency and mistag rates for c and light flavour jets have been measured in the three following η regions $|\eta| = [0 - 0.8]\,, [0.8 - 1.4]$ and $[1.4 - 2.4]$ and seven intervals in $p_T = [20 - 30]\,, [30 - 50]$, $[50 - 100]$, $[100 - 200]$, $[200 - 400]$ and > 400 GeV to obtain two-dimensional efficiency maps. Plots in Figures 6.2 and 6.3 show the b-tagging efficiency and mistag rates for the CSV and CMVA taggers for the 70% b-efficiency working point, as a function of the jet p_T and separately for the three η regions.

Figure 6.2 Efficiency for tagging b-jets in a sample of simulated $t\bar{t}$ events at 13 TeV, as a function of the jet p_T and in the three η regions, for the CSV (blue) and CMVA (red) taggers. Plots are relative the 70% b-efficiency working point.

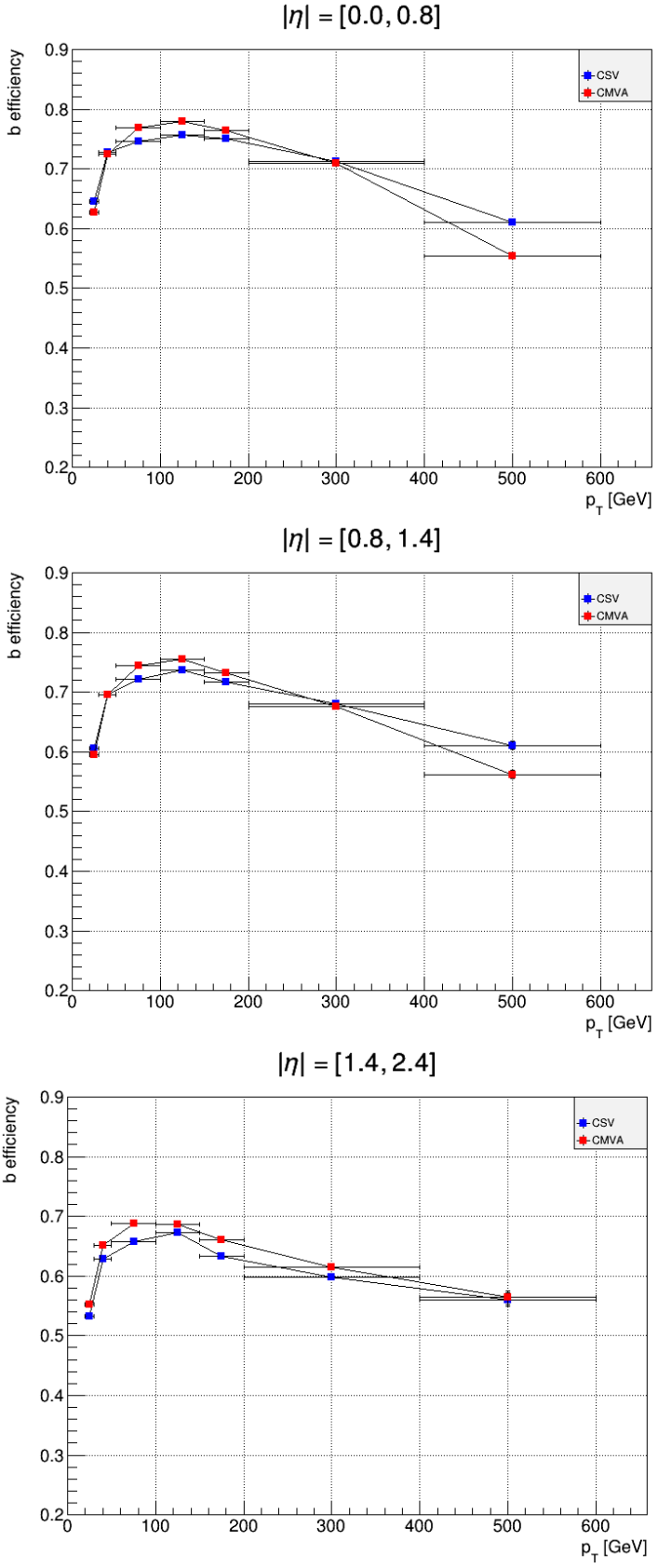
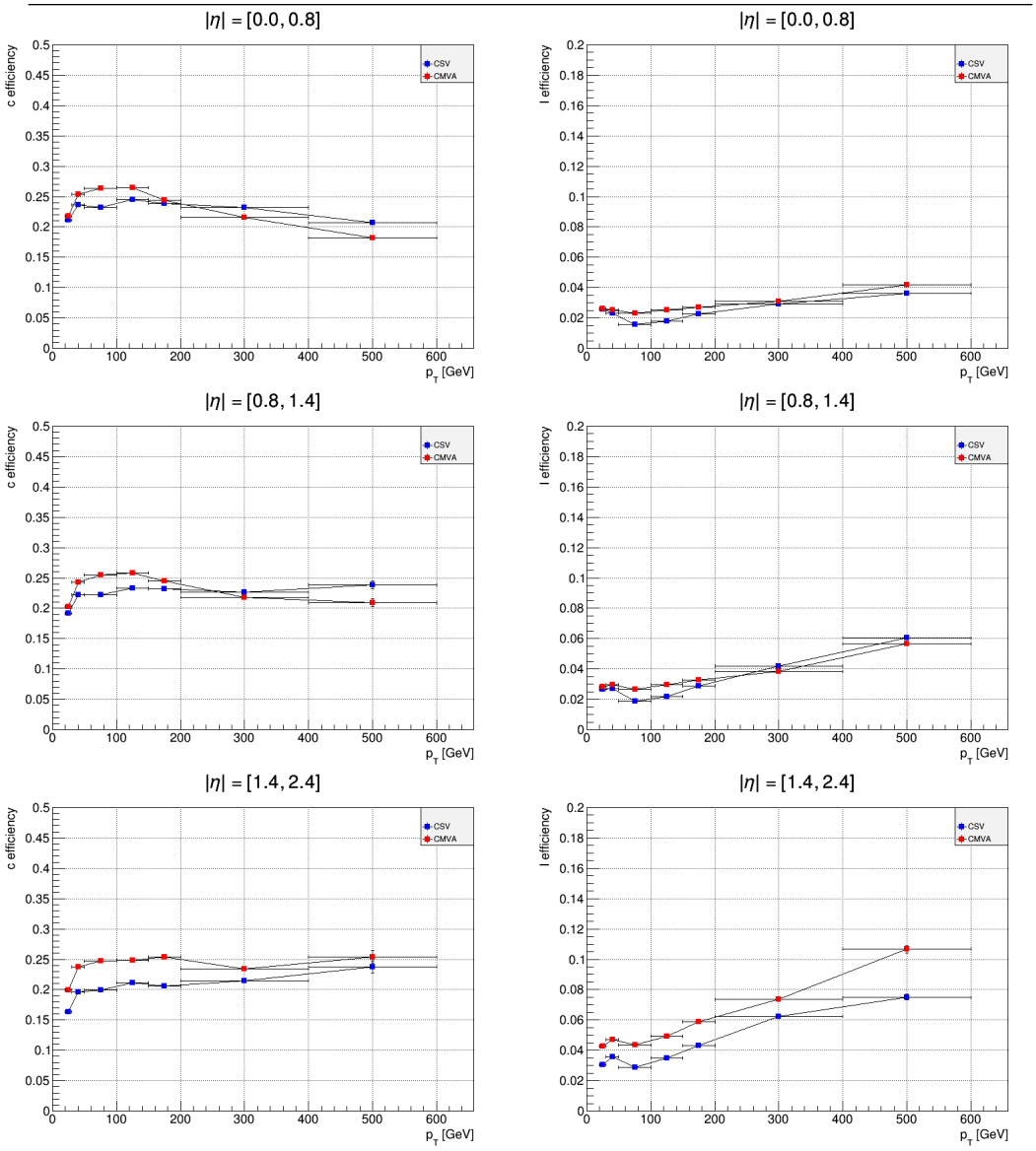


Figure 6.3 Efficiency for tagging jets from c (left) and light (right) flavour jets, as a function of the jet p_T and in the three η regions, for the CSV (blue) and CMVA (red) taggers, in a sample of simulated $t\bar{t}$ events at 13 TeV. Plots are relative the 70% b-efficiency working point.



6.3 Evaluation of the Z(ll)H(bb) Sensitivity

Strategy For each of the scenarios mentioned in Section 6.1, a fully simulated $t\bar{t}$ Monte Carlo sample including detailed simulation of the detector has been used to derive b-tagging efficiencies and mistag rates for c and light jets, as a function of the jet p_T and η, for different operating points. Such efficiencies are then used as weights to be applied to the two jets associated to the Higgs decay, according to their flavour, momentum and direction. In this way, the parametrization of b-tagging efficiencies and corresponding mistag rates are evaluated from events simulated taking into account the different configurations of the upgraded detector and applied to the current signal and background (Drell-Yan and $t\bar{t}$) samples at 13 TeV.

Closure Test The reliability of using a parametrized b-tagging efficiency has been tested on the 13 TeV samples comparing the $S/\sqrt{B}$ ratios obtained using an actual cut on the discriminator or weighting events according to the corresponding efficiency parametrization. The result of such closure test when applying symmetric cuts on the Higgs jets, i.e. when the b-tagging requirement is the same for both the two jets, are presented in Table 6.1, for several working points of the CSV tagger. The discrepancy of the result obtained with the parametrization with respect to the one obtained applying the cut corresponding to the same b-tagging efficiency does not exceed 5% for most of the working points. In other words, the parametrization is closing within about 5%.

Table 6.1 Closure test results for several working points of the CSV algorithm. The $S/\sqrt{B}$ ratios obtained when applying a cut on the discriminator (cut) and when weighting events accordingly to the parametrized efficiency (param.) as a function of jet flavour, p_T and η are compared. Results are for the electron and muon channels combined.

Closure test - Channels combined

Efficiency [%]	CSV discriminator	$S/\sqrt{B}$		Discrepancy [%]
		cut	param.	
50	>0.92	0.727	0.737	1.4
55	>0.90	0.781	0.777	0.5
60	>0.87	0.840	0.817	2.7
65	>0.83	0.870	0.835	4.0
70	>0.77	0.887	0.827	6.7
75	>0.66	0.820	0.777	5.2

Tagger and Working Points Choice In order to evaluate the change in sensitivity of the Z(ll)H(bb) analysis due to b-tagging performances in different detector and pile up conditions, the CSV output has been used as b-tagging discriminator. A loose and a tight working point, corresponding to 75% and 55% nominal efficiency have been chosen for tagging the two jets from the Higgs boson decay. For this operating points, the

parametrization is closing within 5%. The signal region selection is the same as presented in Chapter 5, except for the cut on b-tagging discriminator that is not applied. It is instead replaced by weighting the event as explained in the next Paragraph.

Weighting Procedure In order to apply b-tagging efficiencies evaluated from the available $t\bar{t}$ samples in detector upgrade scenarios, a weighting procedure is performed on signal and background events. First, the efficiencies of tagging a b, c or light jet are measured for the two chosen working points of the CSV algorithm, as a function of the jet p_T and η, in the following bins:
$|\eta| = [0 - 0.8], [0.8 - 1.4]$ and $[1.4 - 2.4]$,
$p_T = [20 - 30], [30 - 50], [50 - 100], [100 - 200], [200 - 400]$ and $[> 400]$ GeV.
As already mentioned, this is done on $t\bar{t}$ samples in detector upgrade scenarios. Then, so obtained efficiencies are applied on the signal and background samples, on an event by event basis: knowing the flavour, p_T and η of the two jets j_1 and j_2 used to reconstruct the Higgs decay, the corresponding efficiencies ϵ_1 and ϵ_2 are given from the measured parametrization. The final weight applied to the event is $w = \epsilon_1 \cdot \epsilon_2$. When using an asymmetric b-tagging requirement, corresponding to two different working points of the b-tagging algorithm, the formula for the weight must be modified. Two different efficiency values corresponding to the looser (ϵ_{loose}) and tighter (ϵ_{tight}) working point are in fact associated to each of the two jets j_1 and j_2. The efficiency corresponding to the requirement that one of the two jets is tightly tagged while the other one is only loosely tagged is given by

$$w = \epsilon_{loose}(j_1) \cdot \epsilon_{tight}(j_2) + \epsilon_{loose}(j_2) \cdot \epsilon_{tight}(j_1) - \epsilon_{tight}(j_1) \cdot \epsilon_{tight}(j_2) \qquad (6.1)$$

The weight assigned to each event is therefore calculated according to Equation 6.1, choosing the working points corresponding to 75% and 55% inclusive b-efficiency for the CSV algorithm.

Results The $S/\sqrt{B}$ ratios, combined over the three Higgs mass bins $m_H = [70 - 100]$, $[100 - 130]$ and $[130 - 160]$ GeV, the Z p_T regions and the Z($\mu\mu$) and Z(ee) channels are listed in Table 6.2 for the different detector upgrade and pile up scenarios, as a function of the recorded integrated luminosity L = 19, 300 or 1000 fb^{-1}. The same values are also plotted in Figure 6.4.

With a new Phase 1 upgraded detector, the improved b-tagging performance benefiting from a new and improved pixel tracker leads to an increase of sensitivity to the Z(ll)H(bb) signal of about 20%, in absence of pile up interactions. When considering an average of 50 interactions per bunch crossing, the loss in sensitivity is only of a few percent, as shown in the left plot of Figure 6.5 that compares the sensitivity of the Z(ll)H(bb) analysis with the actual detector and after the Phase 1 upgrades, for an integrated luminosity of 300 fb^{-1}.

Figure 6.4 Sensitivity of the Z(ll)H(bb) search in several upgrade and pile up scenarios, taking into account the only changes in b-tagging efficiency. The $S/\sqrt{B}$ figure of merit for the electron and muon channels combined is measured for three different values of recorded luminosity. Phase 1 and Phase 2 upgrades are needed to keep the sensitivity almost constant despite the increasing pile up.

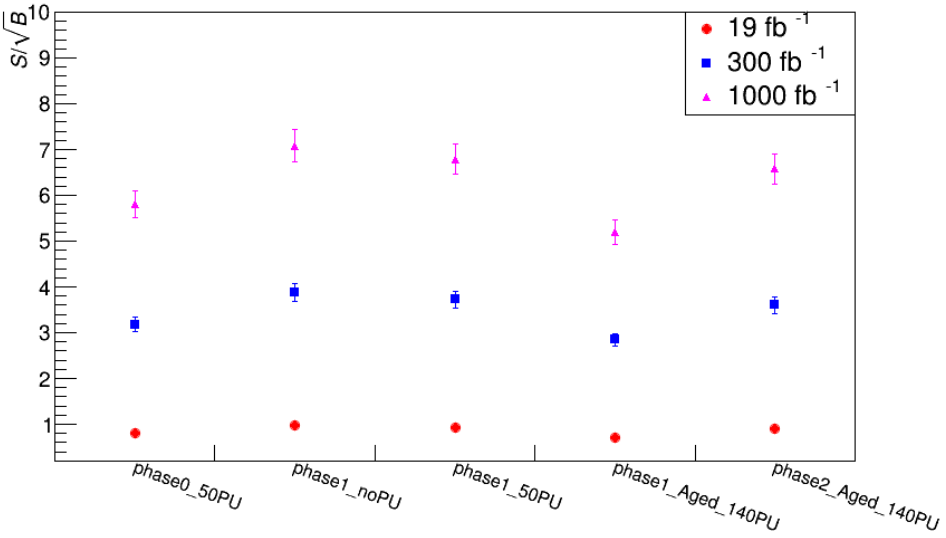


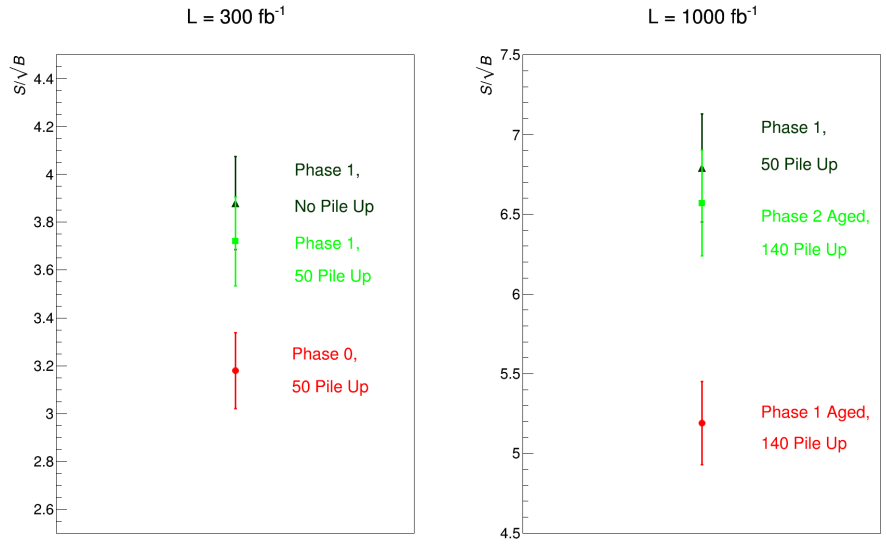
Table 6.2 Sensitivity of the Z(ll)H(bb) search in several upgrade and pile up scenarios, taking into account the only changes in b-tagging efficiency. The $S/\sqrt{B}$ figure of merit for the electron and muon channel combined is measured for three different values of recorded luminosity. An uncertainty of 5% to account for the accuracy of the parametrization evaluated with a closure test is applied.

$S/\sqrt{B}$ - Z(ll) channels combined

Scenario			Integrated Luminosity		
Detector upgrade		Pile Up	19 fb^{-1}	300 fb^{-1}	1000 fb^{-1}
Phase 0		50	0.80 ± 0.04	3.2 ± 0.2	5.8 ± 0.3
Phase 1		0	0.98 ± 0.05	3.9 ± 0.2	7.1 ± 0.4
Phase 1		50	0.94 ± 0.05	3.7 ± 0.2	6.8 ± 0.3
Phase 1	Aged	140	0.72 ± 0.04	2.8 ± 0.1	5.2 ± 0.3
Phase 2	Aged	140	0.91 ± 0.05	3.6 ± 0.2	6.6 ± 0.3

In the high luminosity scenario, an average pile up of 140 interactions is foreseen. In this condition, the performance of the Phase 1 detector after sustaining a delivered luminosity of 1000 fb^{-1} would significantly drop off. Such decrease is mainly due to a reduced hit efficiency of the aged tracker: with a smaller average number of hits associated to a track, discarding fake tracks becomes more challenging. This results in a sensitivity reduction of approximately 25% for the Z(ll)H(bb) signal. The Phase 2 upgrade is therefore needed to keep the detector performance at the same level as after the Phase 1 upgrade even in the more challenging environment of the high luminosity

Figure 6.5 Left: sensitivity of the Z(ll)H(bb) search after the Phase 1 upgrade, compared to the performance of the present detector in a 50 pile up environment, for a luminosity of 300 fb^{-1}. Right: sensitivity of the Z(ll)H(bb) search after the Phase 2 upgrade, compared to the performance of an aged Phase 1 detector in a high pile up environment, for a luminosity of 1000 fb^{-1}. Phase 1 and Phase 2 upgrades are needed to keep the sensitivity almost constant despite the increasing pile up. An uncertainty of 5% corresponding to the closure of the parametrization is taken into account.



operation of the LHC. This target will most likely be achieved for the Z(ll)H(bb) search. As can be seen in the right plot of Figure 6.5, the analysis sensitivity after the Phase 2 upgrades will remain almost unchanged with respect to the realistic Phase 1 scenario, with an average pile up of 50 interactions.

The uncertainty on results presented here is pretty large. Beside the shown uncertainty due to the closure of the b-tagging parametrization, several simplifying assumptions have been made. First, the signal and background cross sections depend on the final center of mass energy that will be reached by the upgraded LHC. Concerning the Z(ll)H(bb) search, predictions done so far assume that the sensitivity depends only on b-tagging efficiency, estimated with the current best algorithm. However, the effect of an optimized signal region selection and of systematic uncertainties in future scenarios can not be estimated yet. Moreover, the contribution to the signal cross section due to ZH production from gluon initiated processes at 13 TeV has not been considered because detailed theory predictions are not available at the time of writing this Thesis, but it is expected to be relatively larger than for the 8 TeV case. Further improvements in sensitivity are likely to arise from the Phase 2 pixel tracker design. In these studies, the the Phase 2 pixel tracker has been assumed to be the same as the Phase 1 pixel detector, as the Phase 2 layout is not yet finalized. However, smaller pixel sizes will more likely be adopted and b-tagging performances will benefit from a higher quality of reconstructed tracks, obtained thanks to a better single hit resolution. Finally, improvements in b-jet identification algorithms might be achieved.

Chapter 7

Conclusions

In this Thesis, the search for the Higgs boson decaying into a pair of bottom quarks with the CMS experiment in the next LHC Runs has been studied. The associated production of the Higgs with a Z boson has been selected, when the latter decays into a pair of same flavour and opposite charge-sign leptons.

In the LHC Run 2 started in May, 2015 protons are being collided at a center of mass energy of 13 TeV. As a first step to prepare the analysis of Run 2 data, some basic kinematic studies have been performed taking into account simulated events of the Z(ll)H(bb) signal and of the main background, that is Drell-Yan production of a Z boson (decaying to a lepton pair) in association with jets. Results of such studies are interpreted making a comparison with the 8 TeV scenario and have been presented in Chapter 5. It has been found that, due to the increased center of mass energy, kinematic distributions of the p_T leading and subleading jets become harder and more central, such differences being more effective for the background than for the signal. Therefore, applying acceptance cuts (requiring at least two central jets and leptons with p_T greater than a given threshold) results in almost the same efficiency for the signal at 13 TeV with respect to the 8 TeV, while the yield of background events passing the selection increases of $\sim$40%. Despite the inclusive production cross sections increase of a factor $\sim$2 for the signal and $\sim$1.7 for the Drell-Yan background with respect to the 8 TeV values, such increase does not directly translate in a higher signal over background ratio. The signal over background ratio is larger at high transverse momentum of the Z boson, similarly to what observed already at 8 TeV. Hence focusing the analysis in the boosted regime is still convenient at 13 TeV.

Following the kinematics studies, a cut-based invariant mass analysis has been performed on simulated events of signal and main backgrounds that have been processed through a full CMS detector simulation. The whole but limited set of events produced for early Run 2 studies have been considered. After some re-optimization with respect to the 8 TeV analysis, the sensitivity of the Z(ll)H(bb) search has been estimated with the $S/\sqrt{B}$ figure of merit.

Combining the electron and muon channels, a sensitivity of $\sim 0.88\sigma$ is expected to be reached with 19 fb^{-1} of collected data at 13 TeV, An optimized choice of the algorithm used to identify b-jets and of the working points improves the analysis sensitivity. The most suitable choice for the Run 2 analysis has been found to be the new CSV algorithm, applying asymmetric cuts on the two Higgs jets. The best cuts correspond to a b-tagging efficiency of $\sim$75% and 50%. With the latter b-tagging selection, the analysis sensitivity increases from 0.88 to 0.92σ, for a luminosity of 19 fb^{-1}. Diboson processes could not be included in the background due to the lack of fully simulated events in Run 2 conditions. A rough estimate suggests that the contribution of the Z(ll)Z(bb) process to the background would decrease the analysis sensitivity of about 5%, from $0.92\ \sigma$ to $0.87\ \sigma$. Moreover, no systematic uncertainties have been taken into account at this stage, as they have not been estimated yet in Run 2 conditions. When assuming the same systematic uncertainty of the Run 1 analysis, the analysis sensitivity would decrease of approximately 9%, from 0.87 to 0.80σ. When comparing the latter result with the one obtained performing the same analysis on 8 TeV data, this translates in a sensitivity increase of approximately 30%, for the same integrated luminosity.

The sensitivity of the Z(ll)H(bb) search has then been studied in several future scenarios, taking into account planned upgrades of the CMS detector and running conditions of the LHC machine. Results of such studies are presented in Chapter 6. Due to the impossibility of producing large enough samples of simulated events for all the future scenarios to be investigated, the changes in b-tagging performance have been evaluated on fully simulated $\mathrm{t\bar{t}}$ events. A two dimensional parametrization of the b-tagging efficiency, as a function of the jet p_T and η, has been derived from the $\mathrm{t\bar{t}}$ samples and extrapolated to estimate the Z(ll)H(bb) sensitivity. The CSV algorithm has been considered for these studies, applying asymmetric cuts corresponding to 75% and 55% b-jet efficiency to tag the two Higgs jets. The accuracy on the analysis sensitivity prediction when using the b-tagging parametrization has been evaluated on the Run 2 samples, comparing results obtained directly applying a b-tagging cut or extrapolating the parametrization derived from $\mathrm{t\bar{t}}$ samples. A good agreement is found, with a discrepancy of only 5%. The most relevant effects of the detector upgrades on the Z(ll)H(bb) sensitivity are provided in Figure 6.5. With a Phase 1 upgraded detector, the improved b-tagging performance benefiting from a new and improved pixel tracker leads to an increase of sensitivity to the Z(ll)H(bb) signal of about 20%, in absence of pile up interactions. When considering an average of 50 interactions per bunch crossing, the loss in sensitivity is only of a few percent. In the high luminosity scenario, an average pile up of 140 interactions is foreseen. In this condition, the performance of the Phase 1 detector after sustaining a delivered luminosity of 1000 fb^{-1} would significantly drop off, mainly due to a reduced hit efficiency of the aged tracker, and would result in a sensitivity decrease of approximately 25% for the Z(ll)H(bb) signal. The Phase 2 upgrade is therefore needed to keep the detector performance at the same level as after the Phase 1 upgrade even in the more challenging environment of the high luminosity operation of the LHC. This target will most likely be achieved for the Z(ll)H(bb) search, whose sensitivity after the Phase 2 upgrades will remain almost unchanged with respect to the realistic Phase 1

scenario, with an average pile up of 50 interactions. The planned upgrades will allow to successfully cope with the increasing pile up and for a luminosity of about 1000 fb^{-1} a sensitivity greater than 5σ for the Z(ll)H(bb) analysis can be reached. Finally, it should be noted that uncertainties on the latter sensitivity predictions are pretty large. Beside the uncertainty due to the closure of the b-tagging parametrization, several simplifying assumptions have been made. First, the signal and background cross sections depend on the final center of mass energy that will be reached by the upgraded LHC. Concerning the Z(ll)H(bb) search, predictions done so far assume that the sensitivity depends only on b-tagging efficiency, estimated with the current best algorithm. However, the effect of an optimized signal region selection, of tighter isolation requirements for leptons that might be necessary in a high pileup scenario as well as systematic uncertainties in future scenarios can not be estimated yet. Moreover, the contribution to the signal cross section due to ZH production from gluon initiated processes at 13 TeV has not been considered because detailed theory predictions are not available at the time of writing this Thesis, but it is expected to be relatively larger than for the 8 TeV case. Further improvements in sensitivity are likely to arise from the Phase 2 pixel tracker design. In these studies, the the Phase 2 pixel tracker has been assumed to be the same as the Phase 1 pixel detector, as the Phase 2 layout is not yet finalized. However, smaller pixel sizes will more likely be adopted and b-tagging performances will benefit from a higher quality of reconstructed tracks, obtained thanks to a better single hit resolution. Finally, improvements in b-jet identification algorithms might be achieved.

This Thesis work also included active participation in the construction of the Phase 1 pixel tracker, that will replace the current CMS pixel detector at the end of 2016. The construction of one half of the third pixel barrel layer is under the responsibility of a consortium of INFN sections. With this respect, the upgrade program currently going on in the Pisa production laboratory mainly includes quality assurance tests of silicon pixel sensors and bare modules, that have been described in Chapter 3. Quality requirements to be satisfied by pixel silicon sensors before dicing are verified measuring their I-V curve. All sensors being used by the Italian consortium for building the BPIX Layer 3 have been tested and only 15% out of the 318 produced sensors did not meet the requirements and have been rejected. The qualification of bare modules is based on more complex tests, aiming at verifying the correct operation of the read out chip and the effectiveness of the bump bonding connections between each pixel on the sensor and its readout cell. Beside testing the electric connection of the pixel to the readout unit, it is also desirable to check the uniformity of the bump solder bumps. In order to achieve this goal without damaging the module, a procedure to verify the bumps height uniformity based on electrical tests has been established. The uniformity of the height of bump connections is found to be uniform at a 2σ level for most of the modules tested so far.

Bibliography

[1] CDF Collaboration, *Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. **74** (1995) 2626, doi:10.1103/PhysRevLett.74.2626.

[2] D0 Collaboration, *Search for High Mass Top Quark Production in $p\overline{p}$ Collisions at $\sqrt{s} = 1.8\ TeV$*, Phys. Rev. Lett. **74** (1995) 2422, doi:10.1103/PhysRevLett.74.2422.

[3] DONUT Collaboration, *Observation of tau neutrino interactions*, Phys. Lett. B **504** (2001) 3 218, doi:10.1016/S0370-2693(01)00307-0.

[4] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321, doi:10.1103/PhysRevLett.13.321.

[5] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508, doi:10.1103/PhysRevLett.13.508.

[6] P. W. Higgs, *Broken Symmetries, Massless Particles and Gauge Fields*, Phys. Lett. **12** (1964) 2 132, doi:10.1016/0031-9163(64)91136-9.

[7] ATLAS Collaboration, *Combined Search for the Standard Model Higgs Boson in pp Collisions at $\sqrt{s} = 7\ TeV$ with the ATLAS Detector*, Phys. Rev. D **86** (2012) 032003, doi:10.1103/PhysRevD.86.032003.

[8] CMS Collaboration, *Combined Results of Searches for the Standard Model Higgs Boson in pp Collisions at $\sqrt{s} = 7\ TeV$*, Phys. Lett. B **710** (2012) 26, doi:10.1016/j.physletb.2012.02.064.

[9] *The 2013 Nobel Prize in Physics - Press Release*, `http://www.nobelprize.org/nobel_prizes/physics/laureates/2013/press.html`.

[10] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264, doi:10.1103/PhysRevLett.19.1264.

[11] S. L. Glashow, *Partial-symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 4 579, doi:10.1016/0029-5582(61)90469-2.

[12] A. Salam, *Weak and Electromagnetic Interactions*, *Proc. of 8th Nobel Symposium*, Almqvist & Wiksell, Stockholm (1968).

[13] Particle Data Group, *Review of Particle Physics*, Chin. Phys. C **38** (2014) 090001, doi:10.1088/1674-1137/38/9/090001.

[14] Wikipedia, *Standard Model*, `http://simple.wikipedia.org/w/index.php?title=Standard_Model&oldid=4920395` (2014).

[15] L. Alvarez-Gaume and J. Ellis, *Eyes on a Prize Particle*, Nat. Phys. **7** (2011) 2, doi:10.1038/nphys1874.

[16] F. Halzen and A.D. Martin, *Quarks and Leptons: an Introductory Course in Modern Particle Physics*, Wiley (1984).

[17] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables* (2011), doi:10.5170/CERN-2011-002.

[18] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, Technical report (2013), Working Group web page: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG, `http://arxiv.org/abs/1307.1347`.

[19] Particle Data Group, *Review of Particle Physics, 2012-2013: Status of Higgs Boson Physics*, Phys. Rev. D **86** (2012) 010001, doi:10.1103/PhysRevD.86.010001.

[20] LEP Working Group for Higgs boson searches, ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, *Search for the Standard Model Higgs Boson at LEP*, Phys. Lett. B **565** (2003) 61, doi:10.1016/S0370-2693(03)00614-2.

[21] CDF Collaboration and D0 Collaboration, *Evidence for a Particle produced in association with Weak Bosons and decaying to a bottom-antibottom Quark Pair in Higgs Boson Searches at the Tevatron*, Phys. Rev. Lett. **109** (2012) 071804, doi:10.1103/PhysRevLett.109.071804.

[22] CDF Collaboration and D0 Collaboration, *Updated Combination of CDF and D0 Searches for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data* (2012), `http://arxiv.org/abs/1207.0449`.

[23] ATLAS Collaboration, *Measurement of Higgs Boson Production in the Diphoton Decay Channel in pp Collisions at center-of-mass Energies of 7 and 8 TeV with the ATLAS Detector*, Phys. Rev. D **90** (2014) 11 112015, doi:10.1103/PhysRevD.90.112015.

[24] CMS Collaboration, *Observation of the Diphoton Decay of the Higgs Boson and Measurement of its Properties*, Eur. Phys. J. C **74** (2014) 3076, doi:10.1140/epjc/s10052-014-3076-z.

[25] ATLAS Collaboration, *Measurements of Higgs Boson Production and Couplings in the four-lepton Channel in pp Collisions at center-of-mass Energies of 7 and 8 TeV with the ATLAS Detector*, Phys. Rev. D **91** (2015) 1 012006, doi:10.1103/PhysRevD.91.012006.

[26] CMS Collaboration, *Measurement of the Properties of a Higgs Boson in the four-lepton Final State*, Phys. Rev. D **89** (2014) 092007, doi:10.1103/PhysRevD.89.092007.

[27] ATLAS Collaboration and CMS Collaboration, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, doi:10.1103/PhysRevLett.114.191803.

[28] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) 08 S08001, doi:10.1088/1748-0221/3/08/S08001.

[29] J. L. Caron, *LHC Layout* (1997), aC Collection. Legacy of AC. Pictures from 1992 to 2002.

[30] J. L. Caron, *Cross Section of LHC Dipole.* (1998), aC Collection. Legacy of AC. Pictures from 1992 to 2002.

[31] J. L. Caron, *Magnetic Field induced by the LHC Dipole's superconducting Coils.* (1998), aC Collection. Legacy of AC. Pictures from 1992 to 2002.

[32] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) 08 S08003, doi:10.1088/1748-0221/3/08/S08003.

[33] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) 08 S08004, doi:10.1088/1748-0221/3/08/S08004.

[34] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) 08 S08005, doi:10.1088/1748-0221/3/08/S08005.

[35] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) 08 S08002, doi:10.1088/1748-0221/3/08/S08002.

[36] TOTEM Collaboration, *The TOTEM Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) 08 S08007, doi:10.1088/1748-0221/3/08/S08007.

[37] LHCf Collaboration, *The LHCf detector at the CERN Large Hadron Collider*, JINST **3** (2008) 08 S08006, doi:10.1088/1748-0221/3/08/S08006.

[38] MoEDAL Collaboration, *The Physics Programme Of The MoEDAL Experiment At The LHC*, Int. J. Mod. Phys. A **29** (2014) 1430050, doi:10.1142/S0217751X14300506.

[39] M. Lamont, *Status of the LHC* **455** (2013) 1 012001, doi:10.1088/1742-6596/455/1/012001.

[40] CMS Collaboration, *CMS Public Luminosity Results*, `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`.

[41] CMS Collaboration, *CMS detector design*, `http://cms.web.cern.ch`.

[42] CMS Collaboration, *Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays*, JINST **5** (2009) arXiv:0910.5530 T03021, doi:10.1088/1748-0221/5/03/T03021.

[43] A. Zabi, *Test Beam Results on the Performance of the CMS Electromagnetic Calorimeter*, Technical report, CERN, Geneva (2006), `http://cds.cern.ch/record/988640`.

[44] CMS Collaboration, *Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data*, JINST **5** (2009) T03012, doi:10.1088/1748-0221/5/03/T03012.

[45] CMS Collaboration, *CMS Physics: Technical Design Report, Vol. 1: Detector Performance and Software*, CERN, Geneva (2006), `http://cds.cern.ch/record/922757`.

[46] CMS Collaboration, *Performance of CMS Muon Reconstruction in pp Collision Events at* $\sqrt{s}$ *= 7 TeV*, JINST **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002.

[47] CMS Collaboration, *Description and Performance of Track and Primary-vertex Reconstruction with the CMS Tracker*, JINST **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009.

[48] Y. Allkofer and others, *Design and Performance of the Silicon Sensors for the CMS Barrel Pixel Detector*, NIM A **584** (2008) 25, doi:10.1016/j.nima.2007.08.151.

[49] CMS Collaboration, *CMS Tracker Technical Design Report* (1998), `https://cdsweb.cern.ch/record/922757`.

[50] CMS Collaboration, *Commissioning and Performance of the CMS Pixel Tracker with Cosmic Ray Muons*, JINST **5** (2009) T03007, doi:10.1088/1748-0221/5/03/T03007.

[51] CMS Collaboration, *Tracker Material Budget Plots*, `https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackerMaterialBudgetplots` (2014).

[52] A. De Cosa and M. Malberti, *Maintenance of the CMS Pixel detector*, `https://cms.web.cern.ch/news/maintenance-cms-pixel-detector`.

[53] L. Rossi, *Common LHC Time Plan: Deliverable: D1.7* (2014), `https://cds.cern.ch/record/1972606`.

[54] CMS Collaboration, *Technical Proposal for the Upgrade of the CMS Detector through 2020* (2011), `https://cds.cern.ch/record/1355706`.

[55] CMS Collaboration, *CMS Technical Design Report for the Pixel Detector Upgrade* (2012), `http://cds.cern.ch/record/1481838`.

[56] CMS Collaboration, *Technical Proposal for the Phase-II Upgrade of the CMS Detector* (2015), https://cds.cern.ch/record/2020886.

[57] CMS Collaboration, *The Tracker Project Technical Design Report*, http://cmsdoc.cern.ch/ftp/TDR/TRACKER/tracker_tdr_files.html (1998), cERN/LHCC 98-6.

[58] CMS Collaboration, *The Performance of the CMS Muon Detector in proton-proton Collisions at $\sqrt{s}$ = 7 TeV at the LHC*, JINST **8** (2013) P11002, doi:10.1088/1748-0221/8/11/P11002.

[59] CMS collaboration, *Performance of CMS Muon Reconstruction in pp Collision Events at $\sqrt{s}$ = 7 TeV*, JINST **7** (2012) 10 P10002, doi:10.1088/1748-0221/7/10/P10002.

[60] CMS Collaboration, *Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s}$ = 8 TeV* (2015), http://arxiv.org/abs/1502.02702.

[61] CMS Collaboration, *Performance of Electron Reconstruction and Selection with the CMS Detector in proton-proton Collisions at $\sqrt{s}$= 8 TeV*, JINST **10** (2015) 06 P06005, doi:10.1088/1748-0221/10/06/P06005.

[62] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t Jet Clustering Algorithm*, JHEP **4** (2008) P63, doi:10.1088/1126-6708/2008/04/063.

[63] CMS Collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET* (2009), cMS Physics Analysis Summary CMS-PFT-09-001, https://cds.cern.ch/record/1194487.

[64] CMS Collaboration, *Performance of Missing Transverse Momentum Reconstruction Algorithms in proton-proton Collisions at $\sqrt{s}$=8 TeV with the CMS Detector* (2013), cMS Physics Analysis Summary CMS-PAS-JME-12-002, https://cds.cern.ch/record/1543527.

[65] CMS Collaboration, *Missing Transverse Energy Performance of the CMS Detector*, JINST **6** (2011) P09001, doi:10.1088/1748-0221/6/09/P09001.

[66] D0 Collaboration, *An improved Determination of the Width of the Top Quark*, Phys. Rev. D **85** (2012) 091104, doi:10.1103/PhysRevD.85.091104.

[67] D0 Collaboration, *Observation of Single Top Quark Production*, http://www-d0.fnal.gov/Run2Physics/top/singletop_observation.

[68] F. Palla A. Rizzi and G. Segneri, *Track Impact Parameter based b-tagging with CMS* (2006), http://cds.cern.ch/record/927385.

[69] CMS Collaboration, *Identification of b-quark Jets with the CMS Experiment*, JINST **8** (2012) P04013, doi:10.1088/1748-0221/8/04/P04013.

[70] CMS Collaboration, *Search for the Standard Model Higgs Boson produced in association with a W or a Z Boson and decaying to Bottom Quarks*, Phys. Rev. D **89** (2013) 012003, doi:10.1103/PhysRevD.89.012003.

[71] ATLAS Collaboration, *Search for the $b\bar{b}$ Decay of the Standard Model Higgs Boson in associated $(W/Z)H$ Production with the ATLAS Detector*, JHEP **1501** (2015) 069, doi:10.1007/JHEP01(2015)069.

[72] CMS Collaboration, *Performance of τ-lepton Reconstruction and Identification in CMS*, JINST **7** (2012) P01001, doi:10.1088/1748-0221/7/01/P01001.

[73] G. Ferrera, M. Grazzini and F. Tramontano, *Associated Higgs-W-Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at NNLO*, Phys. Rev. Lett. **107** (2011) 152003, doi:10.1103/PhysRevLett.107.152003.

[74] M. Ciccolini, A. Denner and S. Dittmaier, *Strong and Electroweak Corrections to the Production of a Higgs Boson + 2 Jets via Weak Interactions at the Large Hadron Collider*, Phys. Rev. Lett. **99** (2007) 161803, doi:10.1103/PhysRevLett.99.161803.

[75] CMS Collaboration, *Summaries of CMS cross section measurements*, `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined` (2015).

[76] B. P. Roe and others, *Boosted Decision Trees, an alternative to Artificial Neural Networks*, NIM A **543** (2007) 577, doi:10.1016/j.nima.2004.12.018.

[77] A. Hoecker and others, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, `http://arxiv.org/abs/physics/0703039`.

[78] CMS Collaboration, *Precise Determination of the Mass of the Higgs Boson and Tests of Compatibility of its Couplings with the Standard Model Predictions using Proton Collisions at 7 and 8 TeV* (2014), `http://arxiv.org/abs/1412.8662`.

[79] CMS Collaboration, *Search for the Standard Model Higgs Boson produced through Vector Boson Fusion and decaying to b bbar*, Sub. to Phys. Rev. D (2015), `http://arxiv.org/abs/1506.01010`.

[80] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD Computations with Parton Shower Simulations: the POWHEG Method*, JHEP **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070.

[81] J. Alwall and others, *MadGraph 5: going beyond*, JHEP **06** (2011) 128, doi:10.1007/JHEP06(2011)128.

[82] M. Bahr, S. Gieseke, M.A. Gigg, D. Grellscheid, K. Hamilton and others, *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639, doi:10.1140/epjc/s10052-008-0798-9.

[83] T. Sjostrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026.

[84] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036.

[85] Geant4 Collaboration, *Geant4 - a Simulation Toolkit*, NIM A **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

List of Figures

List of Tables